

Université de Montréal

**Développement d'approches de contrôle de qualité pour la caractérisation de  
l'immuno-peptidome de cellules infectées par les coronavirus**

Par  
Jérôme Despault-Duquette

Faculté de Médecine  
Département de Microbiologie, Immunologie et Infectiologie

Mémoire présenté en vue de l'obtention du grade de Maître ès Sciences en  
Microbiologie, Immunologie et infectiologie

12/2021

© Jérôme D. Duquette, 2021

Université de Montréal  
Unité académique : Microbiologie, Immunologie et infectiologie

---

*Ce mémoire (ou cette thèse) intitulé(e)*

**Développement d'approches de contrôle de qualité pour la caractérisation de  
l'immuno-peptidome de cellules infectées par les coronavirus**

*Présenté par*  
**Jérôme D. Duquette**

*A été évalué(e) par un jury composé des personnes suivantes*

**Christian Beauséjour**  
Président-rapporteur

**Etienne Caron**  
Directeur de recherche

**Frédéric-Antoine Mallette**  
Membre du jury

## Résumé

La présentation de l'antigène est un mécanisme par lequel les cellules nucléées présentent un court peptide sur la molécule de classe 1 du Complexe Majeur d'Histocompatibilité (CMH-1) codée par les gènes « antigènes d'histocompatibilité humains ». Le terme "immunopeptidomique" est utilisé pour décrire l'ensemble des peptides associés aux molécules du CMH-1. Les cellules T CD8+ patrouillent l'organisme, s'attachent à la molécule CMH-1 par leur récepteur T et détruisent les cellules affichant un peptide atypique. Ce domaine présente un grand intérêt au niveau du traitement des infections virales et dans la conception de vaccins.

Compte tenu que les coronavirus ont été à l'origine de trois épidémies durant les 20 dernières années, et que de multiples souches circulent chez l'humain ainsi que dans le règne animal, il est impératif de développer des vaccins universels qui pourrait prévenir de futurs événement épidémiologiques mondiaux reliés aux coronavirus.

L'immunopeptidomique souffre d'un manque de protocoles normalisés et de contrôle et d'assurance de la qualité des échantillons afin de libérer tout son potentiel dans la recherche biomédicale. Dans le cadre de cette étude, la spectrométrie de masse, la cytométrie de flux et des approches bio-informatiques ont été utilisées pour développer des protocoles de contrôle de qualité pour la caractérisation de l'immunopeptidome de cellules infectées par les coronavirus. Nous avons isolé et analysé l'immunopeptidome de cellules MRC-5 avant et après infection par le coronavirus humain OC-43. En plus d'observer une forte baisse de l'abondance des molécules HLA et de la variété des peptides présentés après l'infection, 9 peptides viraux ont été isolés à partir des molécules du CMH-1. Ces peptides pourraient être utilisés afin de contribuer à formuler un vaccin pan-coronavirus qui élicite une réponse balancée entre la réponse humorale et la réponse cytotoxique.

**Mots-clés :** complexe majeur d'histocompatibilité de classe 1 (CMH-1), antigènes d'histocompatibilité humains (HLA), immunopeptidomique, coronavirus, spectrométrie de masse, expression quantitative de récepteur de surface, QIFIKIT.

## Abstract

Antigen presentation is a mechanism by which nucleated cells present a short peptide on the Major Histocompatibility Complex (MHC-1) class 1 molecule encoded by the “human histocompatibility antigen” genes. The term "immunopeptidomics" is used to describe the set of peptides associated with MHC-1 molecules. CD8+ T cells patrol the body, attach to the MHC-1 molecule through their T receptor and destroy cells displaying an atypical peptide. This field is of great interest in the treatment of viral infections and in vaccine design.

Given that coronaviruses have been responsible for three epidemics in the last 20 years, and that multiple strains circulate in humans and animals, it is imperative to develop universal vaccines that could prevent future global epidemiological events related to coronaviruses.

Immunopeptidomics suffers from a lack of standardized protocols and sample quality control and assurance to unleash its full potential in biomedical research. In this study, mass spectrometry, flow cytometry, and bioinformatics approaches were used to develop quality control protocols for characterizing the immunopeptidome of coronavirus-infected cells. We isolated and analyzed the immunopeptidome of MRC-5 cells before and after infection with human coronavirus OC-43. In addition to observing a strong decrease in the abundance of HLA molecules and in the variety of peptides presented after infection, 9 viral peptides were isolated from MHC-1 molecules.

These peptides could be used to help formulate a pan-coronavirus vaccine that elicits a balanced response between humoral and cytotoxic responses.

**Keywords:** Class-1 Major histocompatibility complex (MHC-1), Human leukocyte antigen (HLA), Coronavirus, immunopeptidomics, mass spectrometry (SM), cell-surface quantitative expression, QIFIKIT.

# Table des matières

Résumé .....	3
Abstract .....	4
Table des matières .....	5
Liste des figures .....	10
Liste des tableaux.....	12
Liste des sigles et abréviations .....	13
Aperçu .....	15

## CHAPITRE 1 - Introduction

1.1 - Le système immunitaire : l'immunité innée et acquise .....	17
• 1.1.1 - Composantes du système adaptatif à médiation cellulaire.....	18
• 1.1.1.2 - Les lymphocytes T .....	19
▪ 1.1.1.2.1 - Les lymphocytes T CD4+.....	20
▪ 1.1.1.2.2 - Les lymphocytes T CD8+.....	21
• 1.1.1.3 Les cellules présentatrices d'antigènes .....	21
1.2 - La présentation de l'antigène .....	22
• 1.2.1 Le complexe majeur d'histocompatibilité .....	22
• 1.2.1.1 - Le complexe majeur d'histocompatibilité de classe 1 .....	22
▪ 1.2.1.2 La voie de présentation classique du CMH-1 .....	24
▪ 1.2.1.3 La voie de présentation croisée des peptides exogènes par le CMH-1 .....	26
• 1.2.2.1 Le complexe majeur d'histocompatibilité de classe 2 .....	26
▪ 1.2.2.2 La présentation antigénique du CMH-2.....	26
1.3 Analyse de l'immunopeptidome .....	29
• 1.3.1 - Les techniques d'analyses de l'immunopeptidome .....	31
• 1.3.1.1 - La quantification des récepteurs du CMH de surface .....	31
• 1.3.1.2 - L'isolation des peptides .....	32
▪ 1.3.1.2.1 - L'élution acide douce .....	32
▪ 1.2.1.2.2 - L'élution par Immuno-affinité .....	34

- 1.3.1.3 - La Spectrométrie de masse et l'identification de l'immunopeptidome ..... 34
  - 1.3.1.3.1 -L'acquisition de données ..... 36
    - 1.3.1.3.1.1 - L'acquisition dépendante de données (DDA)..... 36
    - 1.3.1.3.1.2 - L'acquisition de données ciblées..... 37
    - 1.2.1.3.1.3 - L'acquisition de données indépendante de données (DIA) ..... 38
  - 1.3.1.3.2 - Analyses des données ..... 40
    - 1.3.1.3.2.1 - Traitement des données brutes ..... 41
    - 1.3.1.3.2.2 - Analyses de prédiction de liaison Peptidique ..... 41
    - 1.3.1.3.2.3 - Analyse intégrée des peptides associés au CMH : MHCvizPipe..... 42
- 1.4 - Immunopeptidomique et pandémie de coronavirus ..... 43
- 1.5 - Contexte de recherche ..... 48
  - 1.5.1 - Objectifs de recherche ..... 48

**CHAPITRE 2 – Isolation des peptides du CMH par purification d'immuno-affinité : contrôle de la qualité à partir de lignées cellulaires de références.**

- 2.1 - Lignées de références - contrôle de la qualité des immunopeptidome ..... 50
  - 2.1.1 - Choix des lignées cellulaires appropriés ..... 50
  - 2.1.2 - Lignées cellulaires modèles..... 51
- 2.2 Matériel et méthodes..... 52
- 2.3 - Résultats..... 58
  - 2.3.1 - Analyse quantitative du MHC-1/2 de surface par Immunofluorescence ..... 58
  - 2.3.2 - L'Isolation des peptides du CMH par purification d'immuno-affinité... 59
- 2.4 - Discussion..... 61

- 2.4.1 - Analyses de l'abondance des récepteurs de surface..... 61
- 2.4.2 - Optimisation des protocoles d'isolation des peptides du CMH..... 62

**CHAPITRE 3 – Logiciels pour évaluation rapide de la spécificité et qualité générale des immunopeptidomes identifiés**

3.1 - Problématiques de contrôle de la qualité en immunopeptidomique .....	64
• 3.1.1 - MHCvizPipe : L'avenir de l'analyse d'immunopeptidomique .....	64
3.2 - Méthode : Composantes utilisées dans MHCvizpipe .....	66
• 3.2.1 - NetMHCpan et NetMHCIIpan.....	67
• 3.2.2 - GibbsCluster .....	69
• 3.2.3 - MVP annote les peptides à l'aide de NetMHCpan/IIpan.....	70
• 3.2.4 - MVP génère un rapport HTML détaillé.....	70
• 3.2.5 - Aperçu de l'échantillon .....	70
• 3.2.6 - Résultats d'annotation avec cartes thermiques.....	71
• 3.2.7 - Séquence des motifs de liaisons.....	71
3.3 - Résultats.....	72
• 3.3.1 - Enjeux de recherches reproductibles en immunopeptidomique par spectrométrie de masse .....	72
• 3.3.2 - Reproductibilité intra-utilisateur .....	72
• 3.3.3 - Reproductibilité inter-utilisateur/intra-établissement.....	76
3.4 - Discussion : L'avenir des analyses des ligands du MHC par spectrométrie de masse.....	83

**CHAPITRE 4 – Reproductibilité inter-établissement, application des techniques d'isolation et utilisation des logiciels pour l'analyse de l'immunopeptidome de cellules permissives à l'infection au Coronavirus.**

4.1 - La reproductibilité inter-établissement dans l'étude de l'immunopeptidomique .....	85
4.2 - Matériel et méthodes.....	86

4.3 - Résultats .....	87
• 4.3.1 - Reproductibilité entre différents « Pipeline » d'isolation, acquisition et analyse des données inter-établissement.....	87
• 4.3.2 - La disparité des logiciels d'analyse de données.....	91
4.4 - Discussion : L'harmonisation des données inter-institution est déficiente dans le domaine de l'immunopeptidomique .....	92
4.5 - Modèle d'analyse d'immunopeptidome de cellules susceptibles aux coronavirus : MRC-5 et OC43.....	93
4.6 - Matériel et méthodes .....	96
4.7 – Résultats.....	97
• 4.7.1 - Les cellules MRC-5 affichent une forte expression des molécules du CMH-1 .....	97
• 4.7.2 - L'immunopeptidome des cellules MRC5 naïves.....	100
• 4.7.3 - L'immunopeptidome des cellules MRC5 infectées par virus OC43.....	101
4.8 - Discussion : Les avantages et inconvénients de l'utilisation du modèle MRC-5/OC43 .....	105

## **CHAPITRE 5 – Discussion et perspectives**

5.1 - L'avenir de l'étude du complexe majeur d'histocompatibilité au sein de l'approche de la « Biologie des systèmes » .....	107
• 5.1.2 - Le rôle du CMH dans le cadre des cancers, maladies auto-immunes et des maladies infectieuses .....	109
5.2 - Le manque de contrôle de la qualité des échantillons étudiés en immunopeptidomique.....	111
• 5.2.1 - L'importance de la standardisation des méthodes d'isolation et d'analyses des peptides associés au CMH : collecter, harmoniser et partager les données immunopeptidomiques générées par la communauté.....	112
• 5.2.2 - Stratégies pour l'optimisation de la reproductibilité dans l'étude de l'immunopeptidomique inter-établissements.....	114

5.3 - L'importance de la réponse T dans le contrôle de l'infection à SRAS-CoV2 par vaccination .....	115
5.4 - La prochaine génération en immunopeptidomique : prédire, traiter, faire le suivi et guérir les cancers, maladies auto-immunes et infectieuses .....	118
<b>CONCLUSION</b> .....	119
<b>BIBLIOGRAPHIE</b> .....	121
 <b>ANNEXES</b>	
Annexe 1 - Matériel et horaire optimal pour culture cellulaire à grande échelle..	131
Annexe 2 - Protocole pour la quantification des récepteurs MHC de surface par QIFIKIT <sup>R</sup> .....	134
Annexe 3 - Protocole détaillé de l'isolation des peptides HLA par immuno-affinité .....	138
Annexe 4 - "MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- to Large- Scale Immunopeptidome Data Sets" .....	144
Annexe 5 - Immunopeptidome des cellules MRC5 naïves et infectés par Coronavirus OC-43 .....	182

# Liste des figures

## Chapitre 1

**Figure 1.** Représentation schématique de l'hématopoïèse

**Figure 2.** La maturation des cellules T dans le thymus.

**Figure 3.** Structure des molécules du CMH I et II

**Figure 4.** Voie de présentation classique de l'antigène sur molécule du CMH de classe I

**Figure 5.** Voie de présentation de l'antigène sur la molécule du CMH de classe II

**Figure 6.** Représentation schématique de l'élution à acide douce et l'élution par immuno-affinité.

**Figure 7.** Méthodes d'acquisition des données en spectrométrie de masse

**Figure 8.** Neutralisation des virus SARS-CoV-2, SARS-CoV, WIV-1 et RsShC014 par l'anticorps neutralisant DH1047

## Chapitre 2

**Figure 9.** Abondance de molécules du CMH I et II à la surface de diverses lignées de cultures cellulaire observé avec l'utilisation du QIFIKIT.

**Figure 10.** Schéma de la procédure complète pour l'isolation des peptides du CMH de classe I et II

**Figure 11.** Coloration de gel de Coomassie pour suivre l'efficacité de la liaison des anticorps aux billes de sépharose activées au CNB

**Figure 12.** Immuno-transfert pour le suivi des complexes CMH-peptides après élution acide des billes activées par CNBr couplées à des anticorps.

## Chapitre 3

**Figure 13.** Aperçu du rapport MHCvizPipe : deux répliquats d'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules JY

**Figure 14.** Performance à haute vitesse du logiciel MVP pour le contrôle de la qualité en immunopeptidomique basée sur la spectrométrie de masse

**Figure 15.** Rapport MVP détaillé suite à l'analyse de trois répliquats d'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules JY par le même utilisateur.

**Figure 16.** Identification inter-utilisateurs de peptides du CMH de classe I à partir de cellules JY et EL4

**Figure 17.** Rapport MVP suite à l'analyse de l'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules JY par deux utilisateurs d'un même laboratoire

**Figure 18.** Rapport MVP suite à l'analyse de l'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules EL4 par deux utilisateurs d'un même laboratoire.

## **Chapitre 4**

**Figure 19.** Diagramme Venn illustrant les peptides communs suite à l'analyse de l'immunopeptidome de 90 millions de cellules A549-ACE2/TMPRSS2 effectué par deux laboratoires distincts.

**Figure 20.** Rapport MVP suite à l'analyse de l'immunopeptidome de 90 millions de cellules A549-ACE2/TMPRSS2 effectué par deux laboratoires distincts.

**Figure 21.** Résultats de l'analyse de données brutes d'immunopeptidome de cellules A549-ACE2/TMPRSS2 avant et après infection par SRAS-CoV2 analysés par deux différents moteurs de recherches

**Figure 22.** Résultats par cytométrie de flux de l'expression des molécules du CMH-1 sur des cellules MRC5 avant et après l'infection par coronavirus OC43.

**Figure 23.** Images microscopiques des cellules MRC-5 naïves et infectés par le virus HCoV-OC43

**Figure 24.** Diagrammes à barres illustrant la proportion des peptides isolés dans les échantillons de cellules MRC5 naïf et infecté ayant une forte ou faible affinité pour les allèles du CMH exprimés par cette lignée cellulaire

**Figure 25.** Diagrammes à barres et « Heat-Maps » illustrant la proportion et répartition des peptides isolés dans les échantillons de cellules MRC5 naïfs et infectés ayant une forte, faible, ou sans affinité pour les allèles du CMH exprimés par cette lignée cellulaire

## Listes des tableaux

### **CHAPITRE 1**

**Tableau 1.** Épitopes du SRAS-CoV-2 avec les prédictions HLAthena correspondantes pour les 6 allèles HLA-I exprimés par les cellules A549 et les 3 allèles HLA-I dans les cellules HEK293T.

### **CHAPITRE 4**

**Tableau 2.** Rang percentile des prédictions d'affinités des peptides viraux générés par NETMHCPan avec les allèles correspondants.

## Liste des sigles et abréviations

A.A : Acide aminé

ABC : *Antibody Binding Capacity*

AGC : *Automatic Gain Control*

ALC : *Alpha Cleavage*

B2-M : Beta2-Microglobuline

BF : *Binding Factor*

CD : Cellules dendritiques

CLIP : *Class II Associated Invariant Chain Peptide*

CMH : Complexe majeur d'histocompatibilité

CNBr : *Cyanogen Bromide*

CPAs : Cellules présentatrices d'antigènes

CQ : Contrôle de la qualité

DDA : *Data-Dependant Acquisition*

DIA : *Data-Independent Acquisition*

ERAP : *Endoplasmic Reticulum Associated Amino Peptidase*

HLA : *Human Leucocyte Antigen*

INF : Interféron

KLD : *Kullback-Leibler Distance*

LC-MS/MS - *Liquid Chromatography Tandem Mass Spectrometry*

LF : *Lenght Fraction*

Li : Chaîne Invariante

MFI : *Mean Fluorescence Index*

MVP : *MHCvizPipe*

PFA : Paraformaldéhyde

PRM : *Parallel Reaction Monitoring*

RBD : *Receptor Binding Domain*

RCBs : Récepteur B

RCTs : Récepteur T

RE : Réticulum Endoplasmique

RT : *Room Temperature*

S/MRM : *Selective/Multiple Reaction Monitoring*

SABC : *Specific Antibody Binding Capacity*

SM : Spectrométrie de masse

SWATH-MS: *Sequential Window Acquisition of all Theoretical Mass Spectra*

TAP : *Transporter associated with antigen processing*

TNF : *Tumor Necrosis Factor*

## Aperçu

La science qui se penche sur l'étude de l'ensemble des peptides associés aux molécules du complexe majeur d'histocompatibilité est nommée « l'immunopeptidomique ». Les avancées dans ce domaine sont poussées par de nouvelles technologies dans le domaine de la spectrométrie de masse (SM). La présentation de l'antigène est un processus hautement régulé par lequel les cellules nucléées présentent un peptide de 8 à 12 acides aminés sur la niche de la molécule du CMH de classe 1. Ceci permet au lymphocytes T-CD8+ qui patrouillent l'organisme de s'attacher à la surface des cellules via l'interaction de leur récepteur T (RCTs) avec le CMH-1 afin de déterminer l'état de la cellule. Le complexe majeur d'histocompatibilité de classe 1 porte dans sa niche des peptides issus de l'activité protéolytique du protéasome et divers enzymes intracellulaires. La librairie de cellules T-CD8+ est en mesure de reconnaître les peptides provenant de la dégradation de protéines aberrantes ou de pathogènes intracellulaires dans les cas de cancers, maladies auto-immunes ou dans le cas d'une infection. Les antigènes des leucocytes humains, ou HLA (de l'anglais Human Leucocyte Antigen) A, B et C, sont connus pour être parmi les gènes les plus polymorphiques du génome humain et ils codent pour la molécule du CMH de classe 1. Ils sont aussi associés à de nombreuses pathologies courantes.

Jusqu'à maintenant, des dizaines de milliers d'allèles des gènes HLA ont été décrits et certaines estimations proposent que ce chiffre pourrait éventuellement se situer dans les millions. Le haut taux de degré de variation génétique inter-individus résulte en une différence significative dans le répertoire des peptides présentés entre individus, ce qui fait de l'étude des ligands de cette molécule une tâche ardue.

La spectrométrie de masse est une approche puissante qui peut être utilisée pour étudier les peptides portés par cette molécule clé du système immunitaire. La sensibilité et fiabilité de cet outil fait de cette méthode d'analyse une des approches les plus précises dans l'identification de peptides aberrants ou de provenance exogène. Toutefois, ce domaine souffre de lacunes au niveau de l'harmonisation des protocoles et des données

ainsi qu'un manque de contrôle et d'assurance de la qualité des échantillons. Dans cette expérimentation nous développons des approches de contrôle de qualité pour la caractérisation de l'immunopeptidome de lignées cellulaires infectées par un virus de la famille des coronavirus. La spectrométrie de masse, la cytométrie ainsi que plusieurs approches bio-informatiques seront utilisées pour développer une plateforme d'étude afin d'étudier l'immunopeptidome de cellules infectées par un pathogène.

Les méthodes d'analyse utilisées dans cet ouvrage pourraient aider la conception de vaccins de seconde génération basés sur la présentation des antigènes viraux les plus immunogéniques provenant de plus d'un gène viral, provoquant ainsi une réponse immune incluant un fort engagement de la réponse des cellules T, potentiellement plus efficace au niveau de la longévité et moins dangereux pour certains patients à risque. Peu de méthodes existent pour quantifier l'abondance du CMH-1 à la surface d'une cellule. Dans cette expérimentation, le QIFIKIT a été utilisé pour quantifier l'expression de base de la molécule du CMH-1 sur diverses lignées cellulaires afin de cibler les meilleurs modèles d'étude ainsi que dans le but de déterminer les variations de l'abondance de la molécule d'intérêt avant et après l'infection par un pathogène.

Ce mémoire est composé de 5 chapitres. Le chapitre 1 est une introduction sur le système immunitaire adaptatif ainsi que les différents mécanismes impliqués dans la présentation de l'antigène sur la molécule du CMH-I et II. Ce chapitre contient également une introduction aux techniques d'isolation des peptides du CMH et à l'utilisation de la spectrométrie de masse dans l'étude de l'immunopeptidome. Le deuxième chapitre décrit les résultats d'isolation de peptides à partir de lignées cellulaires de référence servant de contrôle de la qualité. Le troisième chapitre explore les divers logiciels utilisés pour l'évaluation rapide des analyses des peptides du HLA. Le quatrième chapitre applique les méthodes développées pour faire l'analyse de l'immunopeptidome de cellules MRC-5 infectés par le coronavirus OC43 et le dernier chapitre discute et observe les perspectives offertes par les résultats de ce travail.

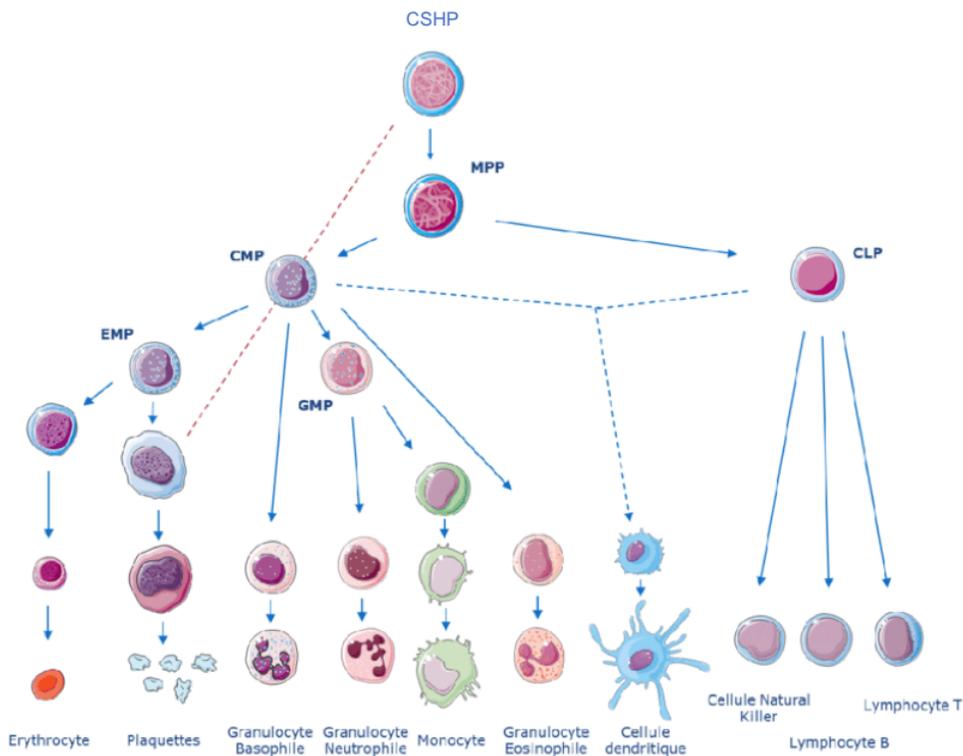
# CHAPITRE 1 - Introduction

## 1.1 Le système immunitaire : l'immunité innée et acquise

Le système immunitaire est considéré comme étant le fruit de modifications et d'incorporations de cascades moléculaires primaires ayant évoluées de manière à provoquer l'émergence de différents types de cellules et organes[1, 2]. Le principal objectif de ce système est d'éradiquer divers dangers pour l'organisme en ciblant et en détruisant la menace au niveau cellulaire. Il exerce ses fonctions en étant capable de reconnaître des modèles moléculaires qui sont exogènes à l'hôte[3]. Le système immunitaire est composé de deux sous-systèmes : L'immunité innée et l'immunité acquise. Le volet inné est considéré comme "non spécifique" car il n'est pas façonné de manière à reconnaître un pathogène précis, mais plutôt conçu pour appliquer une première ligne de défense contre des bactéries, levures, champignons ou virus. Ce système comporte non seulement des éléments physiques et chimiques comme le pH et la fièvre mais aussi une composante cellulaire par la présence de granulocytes: neutrophiles, monocytes, macrophages. Ces différents éléments affectent directement le fonctionnement des pathogènes et activent le système immunitaire inné et adaptatif[4]. L'immunité adaptative ou « acquise » est façonnée de manière plus précise et elle permet à l'hôte, contrairement au système inné, de faire la distinction entre les menaces externes pour ainsi avoir une réponse immune taillée sur mesure contre un pathogène. De manière différente au système inné, cette réponse est lente et dépend de la sélection et de la prolifération clonale de lymphocytes étant en mesure de reconnaître un antigène du non-soi. Ce sous-système est lui aussi divisé en deux par la nature de la réponse immunitaire : la réponse humorale et la réponse à médiation cellulaire ou cytotoxique, respectivement effectuées par les lymphocytes B et lymphocytes T[5]. Ce mémoire focalise son attention sur la réponse adaptative effectuée par la médiation cellulaire.

### 1.1.1 Composantes du système adaptatif

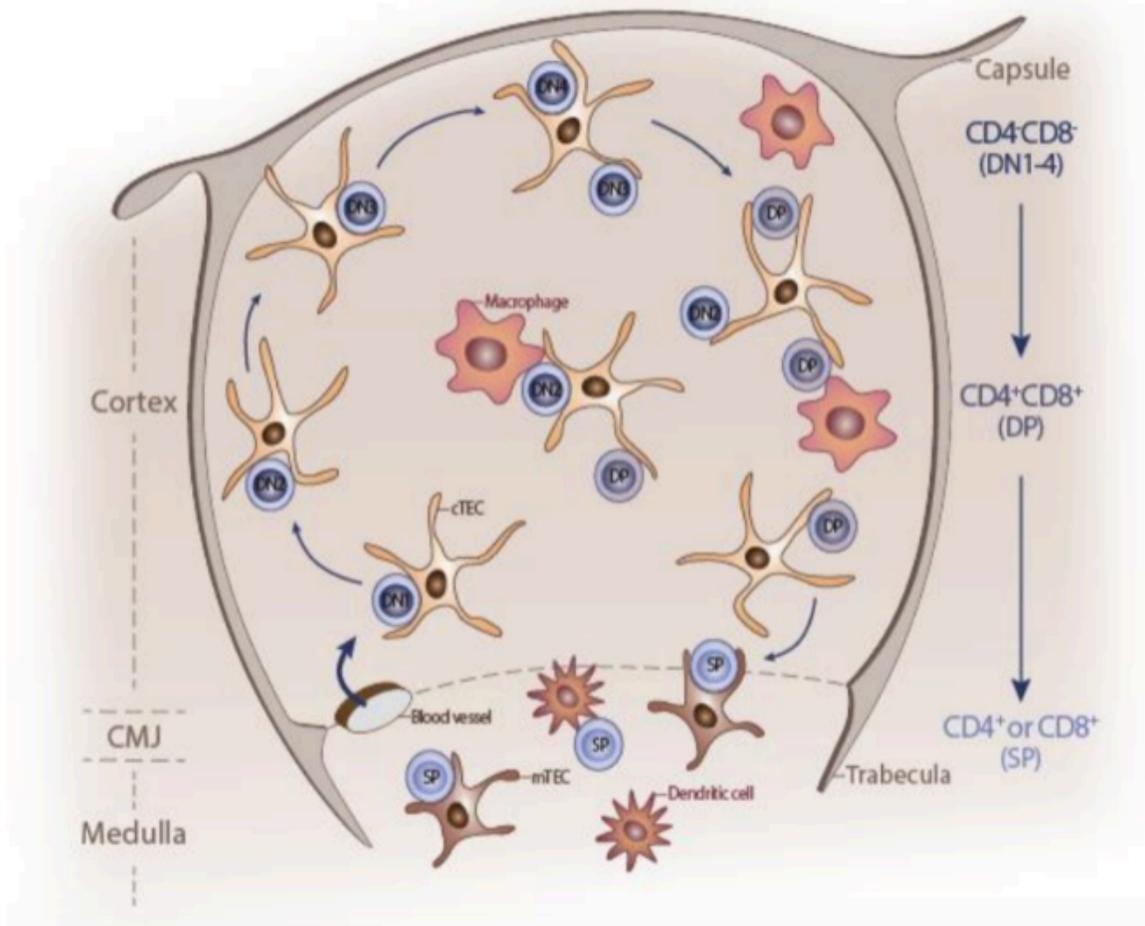
Provenant de la moelle osseuse, les cellules souches pluripotentes hématopoïétiques mûrissent et se différencient progressivement en progéniteurs multipotents jusqu'à atteindre les stades progéniteurs myéloïdes communs, lymphoïdes communs, érythro-mégakaryocytaires ou granulo-monocytaires (**Fig. 1**). Les précurseurs myéloïdes donnent naissance aux cellules érythroïdes, granulocytes et macrophages tandis que les précurseurs lymphoïdes sont à l'origine des cellules T, B, cellules dendritiques (CD) ainsi que les cellules NK [6, 7]. L'activation du système adaptatif dépend des interactions complexes entre les cellules présentatrices d'antigènes, les lymphocytes B, T, et le récepteur du complexe majeur d'histocompatibilité. Les interactions entre les lymphocytes T et B et le CMH se font respectivement par les récepteurs T (RCTs) et récepteur B (RCBs). Les cellules B sont responsables des réponses immunitaires médiées par les anticorps tandis que les RCTs ont la capacité de reconnaître des fragments de protéines présentés à la surface des récepteurs du CMH[8, 9]. Le focus de ce travail étant sur la réponse à médiation cellulaire, on y décrit plus précisément les fonctions des cellules impliquées dans la réponse cytotoxique effectuée par les cellules T.



**Figure 1 :** Schéma de l'hématopoïèse chez l'humain. Cellules souches hématopoïétiques pluripotentes (CSHP), Progéniteur multipotents (MPP), Progéniteur myéloïde commun (CMP), Progéniteur lymphoïde commun (CLP), Progéniteur Erythro-Mégakaryocytaire (EMP), Progéniteur Granulo-Monocytaire (GMP). Adapté de Blank et Karlsson et al et Matherat, G (2015) [10].

### 1.1.1.2 Les lymphocytes T

Suite à leur développement dans la moelle osseuse, les cellules T migrent vers le thymus où elles entreprennent un processus de maturation[6, 7]. Au cours des stades pré-maturation se produit une prolifération où les segments des gènes VDJ sont réarrangés par recombinaison au niveau des RCTs, créant ainsi une vaste librairie de clones de cellules possédant une infinité de récepteurs différents[11, 12]. Au cours des stades de leur croissance, les cellules T subissent un processus appelé éducation thymique, pendant lequel des cellules spécialisées du cortex et de la médulla du thymus façonnent une librairie de cellules T idéale. L'éducation thymique se produit en deux rondes : la sélection positive et la sélection négative (**Fig. 2**). La sélection positive est effectuée dans le cortex du thymus et c'est à cette étape que le thymus sélectionne les clones qui interagissent de manière adéquate avec le CMH de classe I ou II des cellules corticales épithéliales afin de diriger la maturation des lymphocytes double négatifs vers un engagement CD4 ou CD8. Durant la sélection négative, les divers peptides du CMH spécifiques à chaque tissu sont présentés par les cellules du cortex et de la médulla aux cellules T afin d'éliminer les clones qui réagissent trop fortement aux peptides du soi. Ces rondes de sélections donnent naissance à un répertoire de cellules T qui ne réagissent qu'à la présence d'intrus. Les cellules ayant survécu aux deux rondes de sélection échappent le thymus et migrent vers les organes lymphoïdes secondaires où elles sont considérées comme naïves[6, 7, 12]. Elles cherchent par la suite à reconnaître un peptide du non-soi par l'association avec des cellules présentatrices d'antigènes (CPAs)[12, 13]. Il existe deux principaux types de lymphocytes T, les T CD4+ et les TCD8+ qui se différencient dans leurs fonctions mais aussi par la présence de molécules de co-stimulation impliqués dans l'interaction RCTs/CMH[14].



**Figure 2 :** La maturation des cellules T dans le Thymus. Les progéniteurs lymphoïdes communs entrent dans le thymus par la jonction cortico-médullaire (CMJ) à un stade double négatif (DN1-4). Selon l'efficacité de l'interactions avec le CMH de classe I et II et des peptides du soi, les précurseurs mûrent en double positif (DP), ou sont éliminés si trop réactif au soi ou non adaptés à interagir avec le CMH. Les lymphocytes, selon la force de l'interaction avec les récepteurs appropriés, deviennent simple positif (SP) T CD4<sup>+</sup> ou T CD8<sup>+</sup> Adapté de Figueiredo, M et al (2020) [15]

#### 1.1.1.2.1 Les lymphocytes T CD4<sup>+</sup>

Les cellules T CD4<sup>+</sup> stimulent la réponse immunitaire adaptative. Suite à l'association avec une CPA et à l'interaction au sein du RCTs-CD4 et du CMH de classe 2 de la cellule présentatrice, les cellules T CD4<sup>+</sup> s'activent et sécrètent les molécules IL-4 et IL-12 qui favorisent les fonctions de différentes composantes du système immunitaire[16-

18]. Les T CD4+ sont divisés en deux sous-groupes. Les “T-helper 1” ou Th1 et les “T-helper 2” ou Th2. Les Th1 stimulent les fonctions des cellules de la réponse cytotoxique tandis que les Th2 sont responsables d’activer la réponse humorale et la production d’anticorps par les cellules B[19, 20]

#### **1.1.1.2.2 Les lymphocytes T CD8+**

Contrairement aux cellules T CD4+ qui sont restreintes à l’association du CMH de classe 2 des CPA, les cellules T CD8+ peuvent reconnaître le CMH de classe 1 qui est présent à la surface de la grande majorité des cellules nucléées[20, 21]. Ces cellules patrouillent l’organisme et sont activées lorsque se produit une interaction prolongée avec entre le complexe RCTs-CD8 et le CMH-1 d’une cellule cancéreuse ou infectée qui présente un peptide aberrant ne faisant pas partie de la librairie des peptides du soi normalement présentés à la surface de ce récepteur[22]. Suite à cette activation, de nombreuses cascades moléculaires s’activent au niveau de la cellule T, ce qui provoque le relâchement de granzymes et perforines qui attaquent l’intégrité de la cellule anormale[8, 23]. Ce type de cellules a donc un rôle clef dans le contrôle des pathogènes intracellulaires et de la prolifération des cellules cancéreuses. Suite à l’activation du système immunitaire adaptatif se produit une expansion clonale massive des lymphocytes spécifiques à la menace présente[24]. À la suite de l’élimination des cellules ciblées, la grande majorité des clones des lymphocytes T entrent en apoptose. Une faible proportion des cellules effectrices résiste et mûrissent en cellules mémoires[25]. Ce type de cellules peut être activé plus rapidement lors d’une seconde infection avec le même pathogène. C’est ce qu’on nomme la mémoire immune[26, 27].

#### **1.1.1.3 Les cellules présentatrices d’antigènes**

Les macrophages et les cellules dendritiques portent tous un récepteur du CMH de classe 2 et ont la capacité d’interagir avec des lymphocytes T CD4+[23, 28]. Les macrophages, en plus de leur implication dans le système inné, dû à leurs fonctions phagocytaires, ont aussi la capacité de dégrader les pathogènes absorbés et de présenter

un peptide exogène à la surface de sa molécule du CMH de classe 2[28]. Ainsi, ils se comportent en CPA et stimulent l'activation des lymphocytes T CD4+. Les cellules dendritiques agissent en périphérie et décèlent la présence de danger en reconnaissant des patrons moléculaires spécifiques associés à la présence d'inflammation ou de pathogènes bactériens[29]. Ils ont aussi des fonctions phagocytaires qui leur permettent de digérer et de présenter un peptide exogène à la surface de leur récepteur du CMH de classe 2, permettant l'activation de lymphocytes T naïfs[23].

## **1.2 - La présentation de l'antigène.**

Le caractère adaptatif de l'immunité acquise dépend de la présentation de courts peptides antigéniques sur la surface de la membrane cellulaire. Ces fragments de protéines endogènes ou exogènes sont portés dans la niche des molécules du complexe majeur d'histocompatibilité de classe I et II et interagissent respectivement avec les lymphocytes T CD8 + et T CD4+[20, 30, 31].

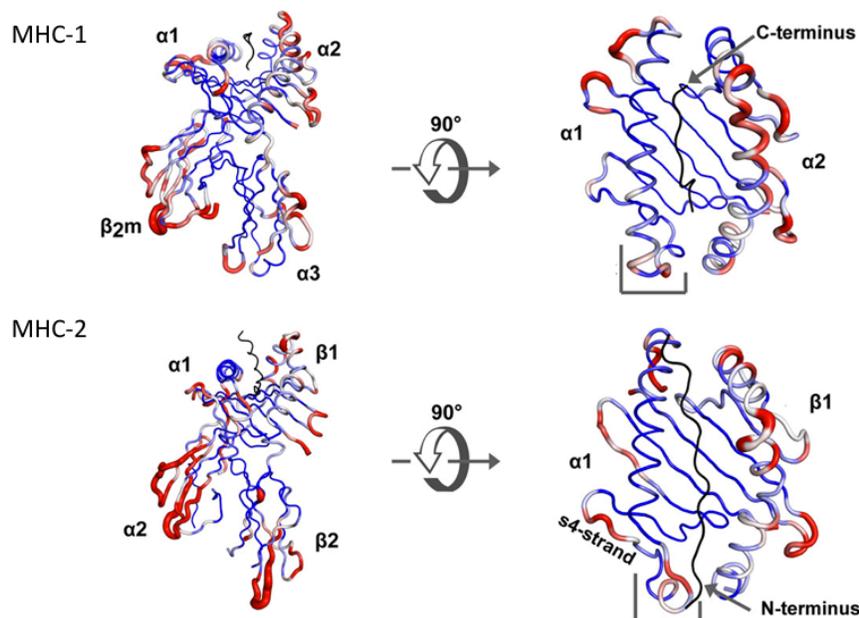
### **1.2.1 - Le complexe majeur d'histocompatibilité.**

#### **1.2.1.1 - Le complexe majeur d'histocompatibilité de classe I**

La molécule du complexe majeur d'histocompatibilité de classe 1 se retrouve à la surface de la quasi-totalité des cellules nucléées[20, 21]. Elle est transmembranaire et comporte dans sa partie externe une niche formée par 8 feuilletts  $\beta$  flanquée de deux hélices  $\alpha$ . C'est dans cette niche que peuvent se loger divers peptides dérivés de la dégradation intracellulaire[32, 33]. Le terme HLA est utilisé pour nommer les molécules du CMH chez l'humain. Les allèles HLA-A, HLA-B et HLA-C sont codés par les gènes du CMH-1a[34, 35]. Les molécules de ce complexe sont connues pour être hautement polymorphiques. Jusqu'à ce jour, 22436 allèles ont été identifiés : 6921 pour l'allèle A, 8181 pour la B et 6779 pour l'allèle C; (<https://www.ebi.ac.uk/ipd/imgt/hla/stats.html>, Aout 2021). Ces variations se situent principalement dans la forme de la niche peptidique et permettent à ces molécules d'interagir avec des peptides ayant des motifs différents selon les individus. Les variations électrochimiques résident principalement dans la

nature des résidus en position P2 et P-Omega (position terminale), qui sont des résidus d’ancrage avec lesquels les peptides s’attachent dans le sillon de liaison du peptide[36]. Par cette grande variabilité dans les gènes du complexe majeur d’histocompatibilité, une population est en mesure de présenter une vaste gamme de peptides et, par le fait même, de produire une immunité de groupe contre un large éventail de pathogènes[37-41].

La molécule du CMH-1 est formée par une chaîne lourde  $\alpha$  de 3 domaines :  $\alpha 1$ ,  $\alpha 2$  et  $\alpha 3$ . C’est à ce trimère que vient se joindre la  $\beta 2$ -microglobuline (B2-M) (**Fig. 3**). La niche peptidique se situe entre les domaines  $\alpha 1$  et  $\alpha 2$  et la liaison dans la membrane est assurée par la portion  $\alpha 3$ . La niche peptidique du CMH-1 est en mesure d’accueillir des peptides de taille variant de 8 à 12 acides aminés. La nature des résidus d’ancrage situés dans le sillon de liaison peptidique définit le motif de liaison du CMH[32, 39, 42, 43]. C’est par la venue de la spectrométrie de masse au début des années 1990 que le séquençage des peptides associés au CMH a permis la découverte des motifs de liaison[44]. De nos jours, plusieurs algorithmes permettent la prédiction de l’affinité des peptides pour les différents allèles du CMH[45-51]. Toutefois, la spectrométrie de masse reste une technique à la fine pointe de la technologie qui peut élucider la présentation de peptides qui peuvent échapper aux prédictions *in-silico* [37, 44, 52-62].



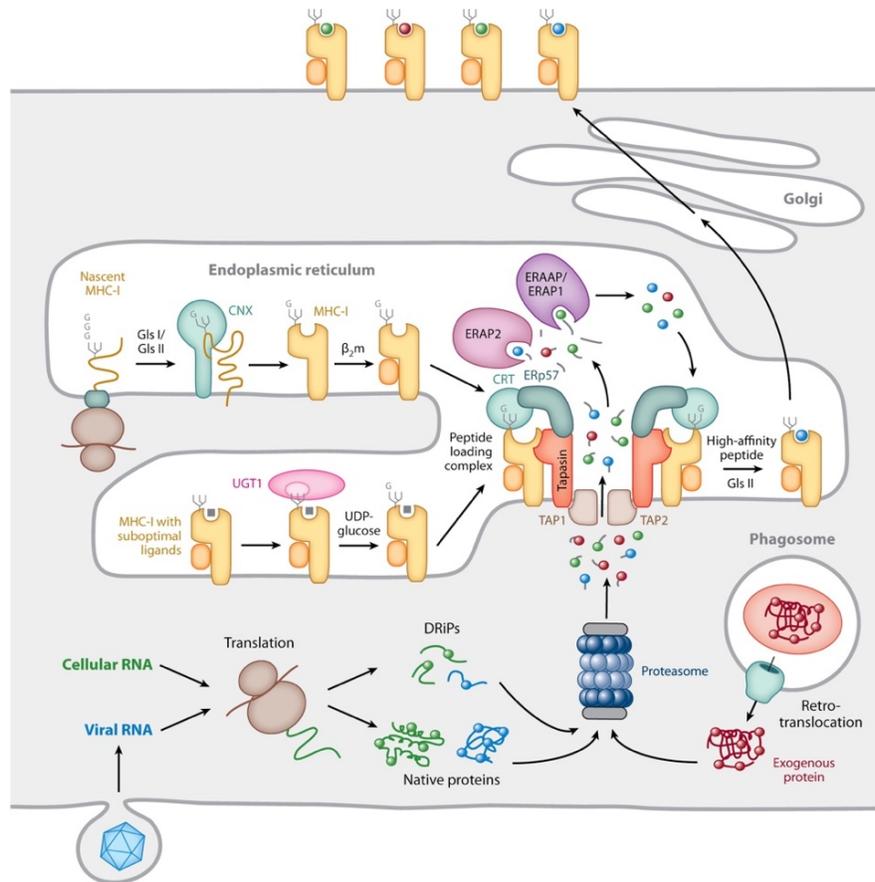
**Figure 3** : Structure du MHC I et II. Représentation schématique d'analyse au rayon X des molécules du CMH I et II vue de côté et de haut. Adapté de Wieczorec, M et al (2017) [63].

### 1.2.1.2 - La voie de présentation classique du CMH-1

La voie de présentation classique des antigènes sur la molécule du CMH-I, comme schématisé à la figure 4, débute par la dégradation des protéines sources par le protéasome[64]. Cet enzyme en forme de tonneau est formé par les sous-unités 20S et 19S pour assembler le protéasome 26S[65, 66]. Les différentes sous-unités forment des anneaux autour d'une chambre où un peptide linéaire peut pénétrer pour y être clivé en plusieurs segments. Les protéines destinées à la dégradation par le protéasome sont marquées par poly-ubiquitination[65]. Elles s'attachent au segment 19S du protéasome avant d'être dépliés, permettant leur entrée dans la cavité du protéasome 20S ou les sous unités catalytiques effectuent le clivage des protéines[66, 67]. L'action de cet enzyme résulte en plusieurs peptides allant jusqu'à environ 20 acides aminés[68]. Il est à noter qu'en présence d'IFN- $\gamma$  et de TNF (*Tumor Necrosis Factor*), trois autres sous-unités catalytiques sont exprimées et viennent remplacer les sous unités B1 B2 et B5 du protéasome constitutif pour former l'immunoprotéasome[69]. Cette variation du protéasome à la particularité de produire des peptides ayant une meilleure affinité avec le CMH-I que ceux produit par le protéasome constitutif en produisant des peptides ayant une meilleure affinité avec les sites d'ancrages principaux de la niche peptidique des molécules du CMH, permettant ainsi une meilleure activation des lymphocytes T CD8+[67, 70]. La présence d'une infection ou l'apparition de tumeurs influence donc fortement la librairie des peptides composant l'immunopeptidome[31, 34, 38, 42, 60, 61, 71]. Les segments de protéines issus de la dégradation protéasomale peuvent ensuite être tronqués par d'autres enzymes cytosoliques avant de se transloquer dans la membrane du réticulum endoplasmique (RE) par l'entremise de la molécule TAP (*Transporter associated with antigen processing*) (**Fig 4**) [72].

Pendant ce temps, la chaîne lourde  $\alpha$  de la molécule du CMH-1 est traduite dans le RE ou elle subit plusieurs étapes de maturation et s'associe à la B-2-microglobulin qui

finit par rejoindre les molécules de Tapasine et ERp57 qui forment le complexe de chargement des peptides situé en proximité immédiate de la molécule TAP (**Fig. 4**)[33]. Avec un apport en ATP, le pore transmembranaire de ce transporteur protéique s'ouvre et permet le passage de segments de protéines qui ont auparavant été clivés par une aminopeptidase cytosolique (ERAP1) à leur extrémité N-terminale pour former des peptides ayant une taille adéquate pour se loger dans la niche du CMH-I[31]. L'attachement des peptides au complexe CMH/B2-M est favorisé par les différentes chaperonnes qui forment le complexe de chargement des peptides : Calréticuline, ERp57 et Tapasine. ERp57/Tapasine serait un complexe clé dans la sélection de peptides possédant un motif favorable à s'associer au sillon du CMH-I (**Fig. 4**)[30, 73]. Le complexe CMH/B2-M/peptide est ainsi stabilisé et peut quitter le complexe de chargement pour être transporté jusqu'à la surface cellulaire par l'entremise de vésicules suite au transport à travers l'appareil de Golgi[30, 33, 72, 74, 75]. Sans infections ou tumeurs, seuls les peptides du "soi" sont associés aux molécules du CMH-I. Ils tiennent de nombreux rôles, notamment dans le développement des cellules T et ils sont aussi impliqués dans de nombreuses pathologies ou phénomènes de rejets de greffes[20, 76].



Blum JS, et al. 2013.  
 Annu. Rev. Immunol. 31:443–73

**Figure 4** : Voie de présentation classique de l'antigène sur la molécule du CMH de classe I. Calréticuline (CRT), Calnexine (CNX). Adapté de Blum et al (2013) [33].

### 1.2.1.3 - La voie de présentation croisée des peptides exogènes par le CMH-1

En plus de la voie classique des peptides endogènes, certaines CPA qui expriment également des molécules du CMH-1, dont les cellules dendritiques et les macrophages, sont en mesure de porter des peptides exogènes à la surface de leur CMH-I via la voie de présentation croisée[77, 78]. Ayant des similitudes avec la voie de présentation des peptides du CMH-II, la présentation croisée peut se produire par un parcours vacuolaire ou cytosolique[77]. Le parcours vacuolaire est initié par l'endocytose d'un antigène extracellulaire, le plus fréquemment par des cellules dendritiques[79]. La vésicule formé accroît sa concentration en acide et déclenche la dégradation des protéines en peptides

par des protéases lysosomales[80]. Il n'est pas encore tout à fait clair si les peptides formés de cette manière sont exportés au RE pour y être chargé sur des molécules de CMH-I naissante ou si elles se lient à des molécules recyclées avant d'être sécrétés à la surface de la cellule par la voie de sécrétion de l'appareil de golgi[81]. De manière similaire à la voie vacuolaire, les antigènes qui emploient la voie cytosolique sont endocytés par des CD avant d'être relâchés dans leur cytoplasme pour être dégradés par le protéasome et suivre la voie de présentation classique[81, 82].

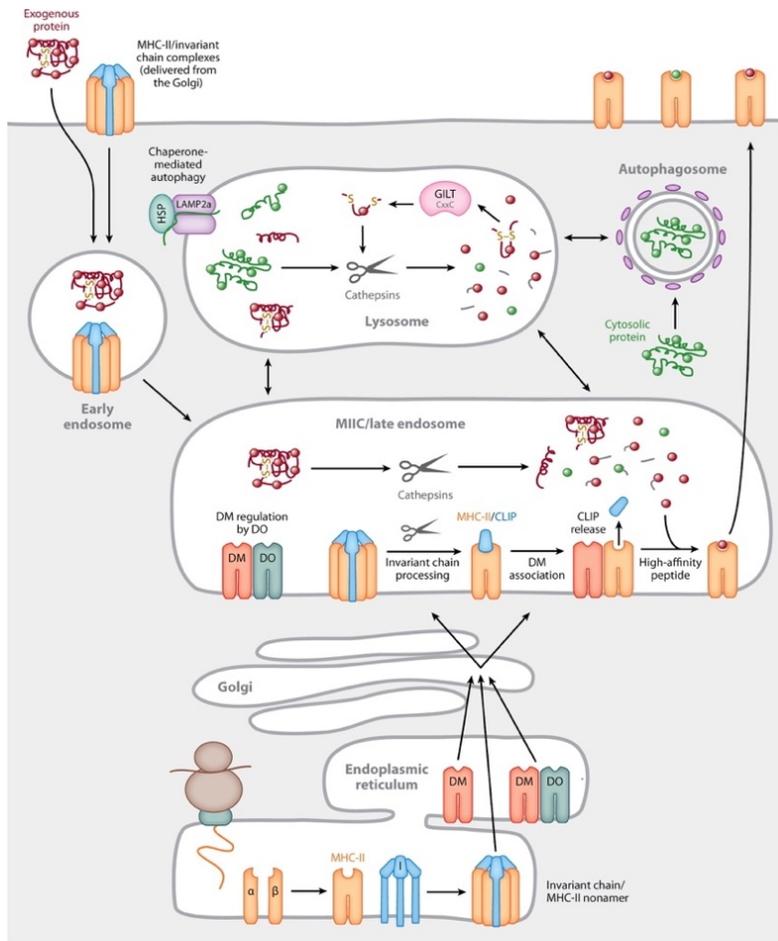
### **1.2.2.1 - Le complexe majeur d'histocompatibilité de classe II**

Les molécules du CMH de classe II se retrouvent exclusivement à la surface de la membrane des cellules présentatrices d'antigènes : Les macrophages, les cellules dendritiques et les cellules B[83]. Cette molécule a comme fonction de présenter des peptides exogènes dans le but d'activer des lymphocytes T CD4+ naïfs[83]. Les peptides du CMH-2 sont dérivés de la dégradation de pathogènes, de cellules infectées ou aberrantes. Les molécules conventionnelles du CMH de classe II (CMH-IIa) sont codées par un groupe de gènes situés sur le chromosome 6 qui codent pour les allèles HLA-DPA, DPB, DQA, DQB, DRA et DRB[84]. La molécule du CMH de classe 2 est composée d'une chaîne  $\alpha$  et d'une chaîne  $\beta$  ayant une portion cytosolique et un domaine externe (**Fig. 3**). La portion externe est composée par les domaines 1 et 2 des chaînes  $\alpha$  et  $\beta$ . De manière similaire à la molécule du CMH de classe I, le CMH-II comporte des régions polymorphiques. Toutefois, la niche peptidique de cette molécule permet la liaison de peptides plus long, allant de 10 à 25 acides aminés grâce à une ouverture aux extrémités du sillon[54]. Cette molécule comporte également des motifs de liaison au niveau des résidus d'ancrages et s'associent de manière préférentielle à certains peptides selon leur propriété électrochimique[83, 85, 86].

### **1.2.2.2 - La présentation antigénique du CMH-II**

La voie de présentation des peptides du CMH-II permet la présentation de peptides exogènes au lymphocytes T CD4+ qui dictent principalement la réponse

humorale mais aussi la réponse cellulaire[31]. Les peptides endocytés par les CPAs qui finissent leur cheminement dans la niche des molécules du CMH-II proviennent généralement de protéines exogènes issus de la dégradation de pathogènes, de cellules infectées ou de cellules cancéreuses. Ces peptides font partie du non-soi[23, 33]. Les chaînes  $\alpha$  et  $\beta$  de la molécule du CMH-II naissent au niveau du RE. Avant la liaison avec un peptide, la présence de la chaîne invariante (Li) stabilise la niche peptidique formée par la jonction des domaines  $\alpha 1$  et  $\beta 1$ . Les complexes MHC-II/Li sont ensuite transportés vers des compartiments de traitement des antigènes dans lesquels une portion de la chaîne Li est dégradée par des cathepsines S et L laissant ainsi dans la niche peptidique un fragment nommé CLIP (*Class II Associated Invariant Chain Peptide*). L'activité de HLA-DM libère le fragment CLIP et permet la liaison d'un peptide exogène à la molécule du CMH-II. Le complexe CMH-II-peptide est transporté jusqu'à la surface de la cellule afin d'interagir avec des lymphocytes T CD4+ (**Fig. 5**) [30, 83].



Blum JS, et al. 2013. Annu. Rev. Immunol. 31:443–73

**Figure 5** - Voie de présentation de l'antigène sur la molécule du CMH de classe II.

Adapté de Blum JS et al (2013) [33]

### 1.3 - Analyse de l'immunopeptidome

La venue de la spectrométrie de masse dans le domaine de l'immunologie a permis de faire les premiers pas afin de définir la nature et l'abondance des peptides liés aux molécules du CMH et de comprendre les fondements du système immunitaire[44]. Ces avancées ont permis l'identification de peptides tumoraux ou viraux en cas de cancers et d'infections[41, 56, 87]. La perception du « soi » par les cellules T est un phénomène qui possède encore de nombreux mystères. Toutefois, plusieurs applications thérapeutiques tirent déjà profit des mécanismes de reconnaissance des cellules T, notamment en immunothérapie contre les maladies auto-immunes, les maladies infectieuses et les

cancers[88, 89]. Une autre avenue qui sera propulsée par les avancés dans ce domaine sera le développement de vaccins de nouvelle génération. L'identification des épitopes viraux observés sur les molécules du CMH-1 dans des échantillons biologiques permettront de concevoir des vaccins plus efficaces en identifiant les séquences virales les plus pertinentes par : (1) leur immunogénicité, (2) leur capacité à interagir avec les molécules du CMH-1 d'une plus grande partie de la population, (3) par leur faible taux de mutations et (4) par la présence de séquence homologues retrouvés chez plusieurs membres de la famille des coronavirus pour le cas des vaccins pan-coronavirus.[60, 90-92]

Les progrès technologiques dans le domaine de la spectrométrie de masse rendront possibles, dans un futur à moyen terme, les études de grandes cohortes en immuno-peptidomique et, en fin de compte, permettront l'association de variations génétiques au niveau des molécules HLA qui fourniront des informations sur la susceptibilité et la résistance aux maladies de certains individus ou groupes de personnes possédant un allèle distinct du CMH[93, 94].

De plus, l'étude de l'immuno-peptidomique sera indispensable dans la guerre constante contre les agents pathogènes à laquelle la race humaine sera confrontée dans les prochaines décennies. Les vaccins « *epitope-based* », basés sur les épitopes les plus immunogènes, présentent un intérêt pour la science car ces vaccins offrent une efficacité thérapeutique maximale avec des effets secondaires minimaux[90, 91, 95].

Toutefois, ce domaine souffre de lacunes au niveau de d'harmonisation des protocoles et des données ainsi qu'un manque de contrôle et d'assurance de la qualité des échantillons. Les méthodes d'analyses décrites dans cet ouvrage permettent l'analyse d'immuno-peptidome d'échantillons cellulaire infectés par un membre de la famille des coronavirus. Le but de ce travail est le développement d'approches de contrôle de qualité pour la caractérisation de l'immuno-peptidome de cellules infectées par un pathogène dans l'optique d'identifier des épitopes viraux présentés à la surface des molécules du CMH-1.

L'avenir de l'immunopeptidomique repose dans la capacité à étudier de larges cohortes et représente un grand défi au niveau des procédures d'échantillonnages. Le développement d'approches permettant une isolation des peptides et une analyse plus rapide et harmonieuse des échantillons entre les institutions pourrait propulser l'utilisation de l'immunopeptidomique dans cette ère de la médecine personnalisée[60, 96-98]. Comme ce mémoire se penche principalement sur l'étude du CMH-I, le terme immunopeptidome se référera à l'immunopeptidome du CMH-I.

### **1.3.1 - Les techniques d'analyse de l'immunopeptidome du CMH-1.**

#### **1.3.1.1 - La quantification des récepteurs du CMH de surface**

Cet ouvrage s'est fortement penché sur l'analyse de l'abondance des molécules du CMH à la surface des cellules de plusieurs lignées de cultures cellulaires. Comme les expériences en SM requièrent des échantillons d'une taille considérable, il est primordial d'utiliser des lignées cellulaires qui expriment fortement les molécules de CMH à leur surface afin de s'assurer de pouvoir en retirer un échantillon biologique adéquat lors de l'isolation des peptides du CMH par immuno-affinité. La quantification de l'abondance de surface cellulaire des molécules du CMH a été rendue possible par un test d'immunofluorescence indirecte[99]. Le QIFIKIT contient une série de 5 types billes de 10 µm de diamètre recouvertes d'une quantité croissantes et définies d'anticorps primaire de souris qui peuvent lier un anticorps secondaire conjugué à un fluorochrome de notre choix[99]. Le kit est utilisé conjointement à un échantillon cellulaire dont les molécules du CMH sont marqués par un anticorps primaire et un anticorps secondaire portant un fluorochrome (le même qu'utilisé avec la série de billes). Les échantillons sont analysés par cytométrie. L'intensité de la fluorescence de la série de bille permet de créer une courbe d'équation linéaire qui permet ensuite d'extrapoler et d'estimer de manière précise la quantité de récepteurs du CMH au niveau de la surface de l'échantillon cellulaire analysé[99, 100]. L'analyse de ce paramètre permet non seulement de choisir des lignées optimales aux recherches précises dans les différentes avenues en immunopeptidomique mais elle permet aussi d'avoir une idée générale de la quantité de

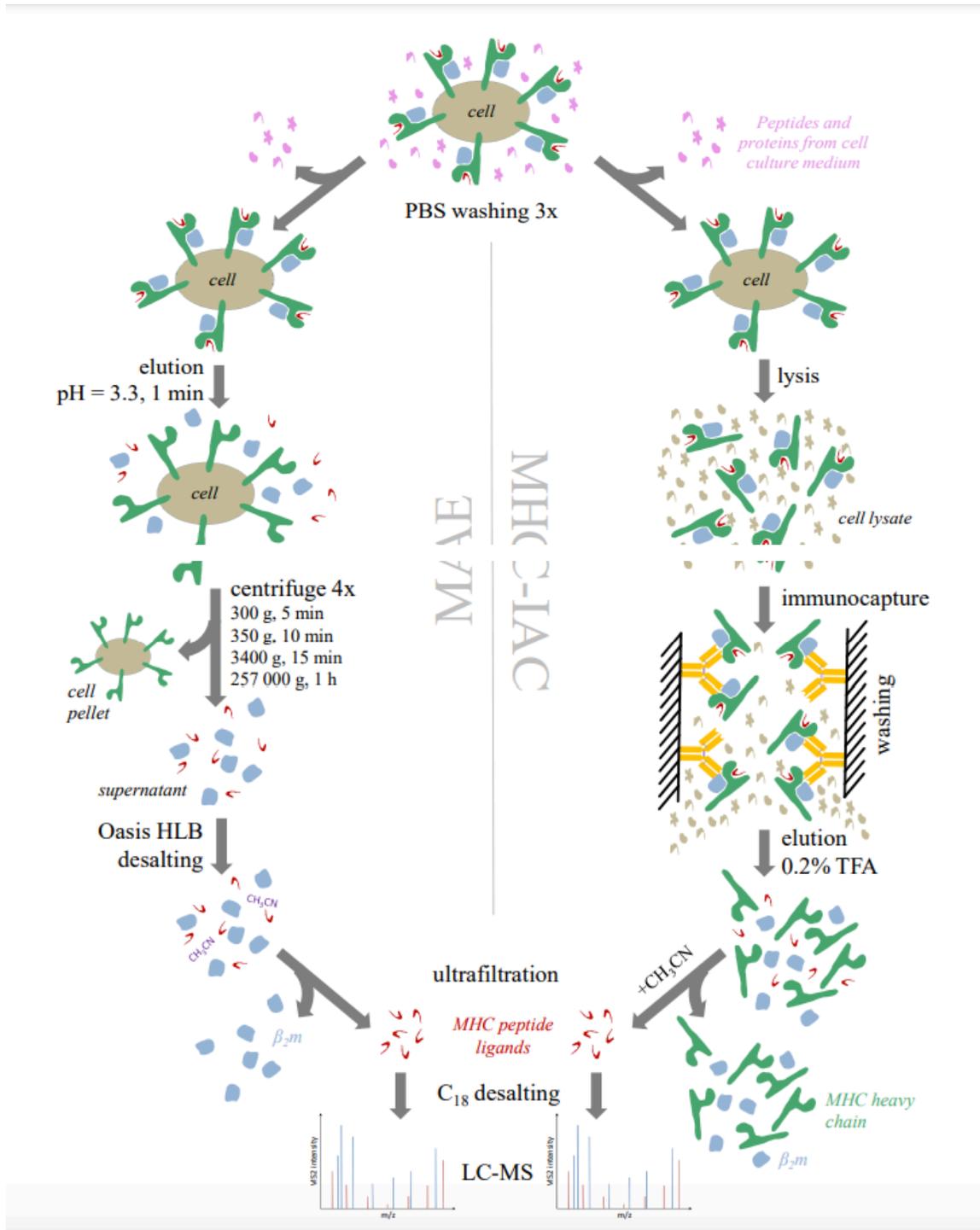
peptides élués qui devraient être présents dans les étapes suivantes des analyses immuno-peptidomiques, donnant ainsi des informations sur le contrôle de la qualité (CQ) des isolations des immuno-peptidomes.

### **1.3.1.2 - L'isolation des peptides.**

L'isolation des peptides du CMH et la préparation qui en découle représentent le "tendon d'Achille" du domaine de l'immuno-peptidomique. Contrairement à de nombreux domaines où les avancées technologiques ont permis la réalisation d'études de larges cohortes, les coûts importants, les manipulations laborieuses et les grands besoins en main-d'œuvre font que cette discipline a pris du retard malgré que son impact potentiel en clinique soit bien reconnu[60, 96, 101].

#### **1.3.1.2.1 - L'élution acide douce.**

Cette méthode est une des plus anciennes méthodes décrites pour isoler les peptides du CMH et, contrairement aux nouvelles méthodes, ne nécessite pas un grand échantillon cellulaire (**Fig. 6**) [102]. Cette technique consiste à traiter des cellules vivantes en suspension avec un tampon citrate. À la suite d'un tel traitement, la molécule de  $\beta$ 2-microglobuline liée de manière non covalente à la chaîne lourde du CMH I se dissocie, déstabilisant la structure de la molécule entière. Ceci réduit la capacité de fixation des peptides des complexes HLA-A, HLA-B et HLA-C, entraînant la perte des peptides associés aux molécules du CMH de classe I sans perturber les interactions entre les molécules du CMH II et leurs peptides car ce complexe ne comprend pas la présence de la molécules B2-M. Malgré le fort rendement de cette méthode, elle est aussi reconnue pour avoir le désavantage d'inclure de nombreux contaminants qui n'ont pas d'affinités pour les molécules HLA. Les contaminants peuvent composer plus de la moitié du matériel protéique isolé à partir de cette méthode[103]. Elle comprend aussi de nombreux désavantages dû au fait qu'elle manque de spécificité et qu'elle est restreinte à être utilisée sur un échantillon cellulaire vivant et en suspension, rendant ainsi impossible l'analyse de tissus qui requièrent des conditions de cultures spécifiques[101].



**Figure 6** – Représentation schématique de l'élué à acide douce et l'élué par immuno-affinité. Adapté de Sturm, T et al (2021) [104]

#### **1.3.1.2.2 - L'élution par immuno-affinité**

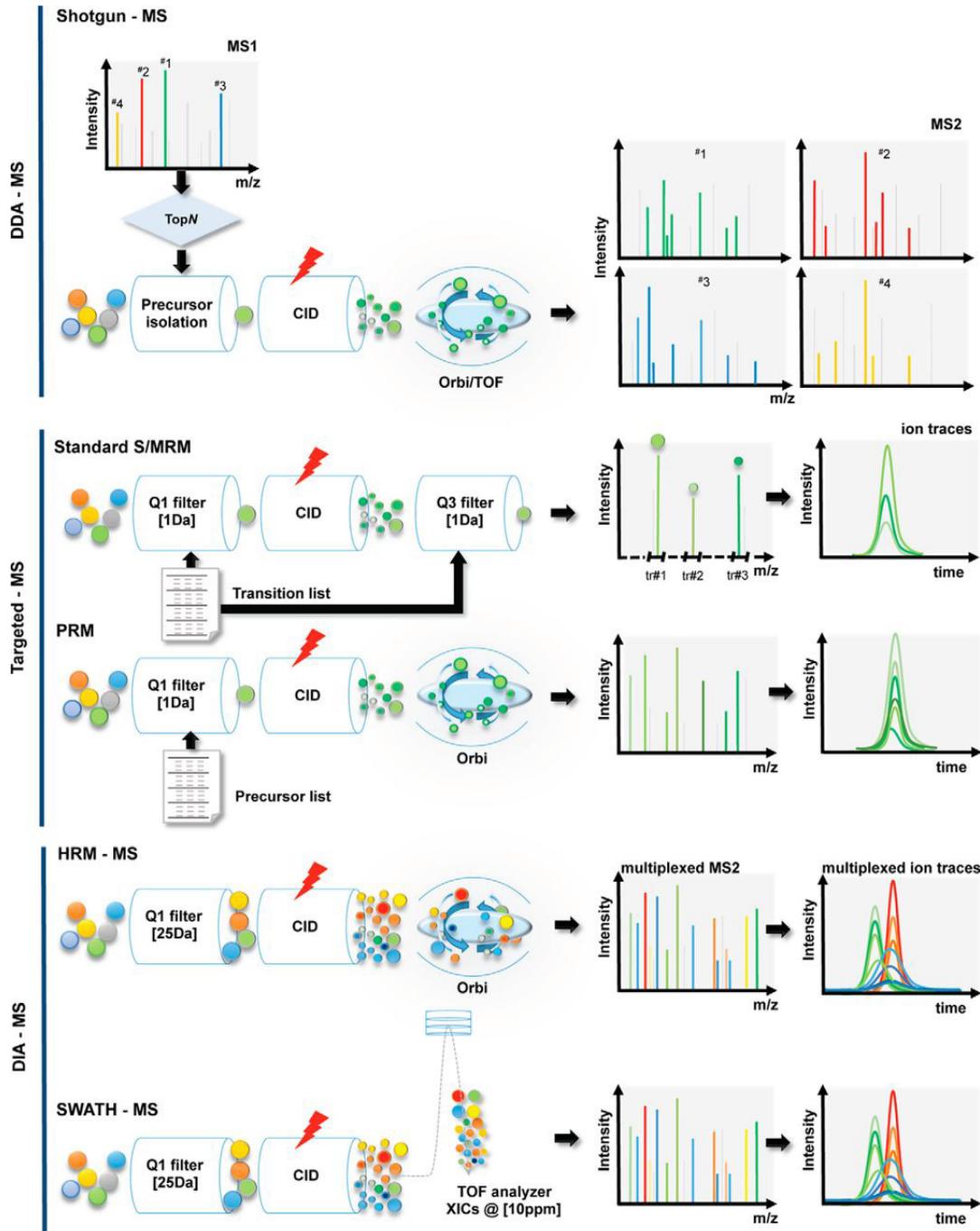
Contrairement à la technique précédente, l'élution par immuno-affinité à l'avantage d'avoir une haute spécificité dans le processus d'extraction des peptides et une plus grande flexibilité dans la nature des échantillons (**Fig. 6**). Un lysat de lignée cellulaire, un tissu homogénéisé, ou un échantillon de fluide biologique est d'abord traité et lysé avec un détergent non dénaturant avant d'être incubé avec des anticorps spécifiques du CMH liés à des résines polymères ou des particules magnétiques qui agissent en tant que support solide. Après purification des substances non spécifiquement liées, les molécules du CMH ainsi que les peptides associés sont élués avec une solution ACN 28%/TFA 0,1%.

Malgré que l'utilisation d'un détergent apporte une présence de contaminants, jusqu'à 90 % des peptides immuno-purifiés en utilisant cette méthode sont spécifiques des molécules du CMH[101]. Un des inconvénients de cette technique est que l'immunopurification par affinité est à l'origine de pertes importantes lors de la manipulation des échantillons. Certaines évidences montrent que jusqu'à 99% des peptides peuvent être perdues lors des différentes étapes de l'immunopurification par affinité. De plus, ces méthodes nécessitent de grandes quantités d'anticorps et de matériel biologique, rendant les études à grande échelle coûteuses et presque inexistantes jusqu'à ce jour. Toutes ces observations prouvent qu'il est impératif de développer de nouvelles méthodologies plus efficaces pour l'isolement et l'évaluation de la qualité générale des échantillonnages. Malgré tout, cette méthode reste la méthode la plus spécifique pour étudier les ligands des molécules HLA[54, 105].

#### **1.3.1.3 - La spectrométrie de masse et l'identification de l'immunopeptidome**

Due à son haut taux de polymorphisme au niveau de la population, à la spécificité des sites de liaison aux peptides et à la présence de multiples allotypes par individus, l'études des ligands des molécules du CMH est une tâche ardue[54, 106]. Au cours de la dernière décennie, les progrès réalisés dans le domaine de la protéomique basée sur la spectrométrie de masse ont grandement amélioré notre capacité à identifier et à quantifier les peptides associés au CMH. Trois types de méthodes d'acquisition de

données en SM sont maintenant disponibles pour l'analyse des peptides associés au CMH : L'analyse des peptides associés au CMH dépendante des données ou DDA (*Data-Dependent Acquisition*), l'acquisition de données ciblées (*Targeted MS*), et l'acquisition indépendante des données ou DIA (*Data-Independent acquisition*) (**Fig. 7**) [54].



**Figure 7 :** Configurations de spectromètres de masse utilisées pour l'analyse des peptides associés au CMH. Toutes les configurations comprennent l'isolation et la fragmentation des ions précurseurs et l'enregistrement du rapport masse/charge des fragments d'ions MS2 dans un analyseur de masse. (Panneau supérieur) En mode DDA, seuls les ions précurseurs les plus abondants (TopN) par balayage MS1 sont sélectionnés pour une isolation et une fragmentation consécutive. Des appareils Orbitrap ou TOF sont utilisés pour l'acquisition à haut débit des spectres MS1 et MS2. (Panneau du milieu) Dans la SM ciblée, le triple quadripôle est utilisé pour cribler une liste de transitions (tr), c'est-à-dire des paires d'ions précurseur-fragment. Les transitions sont suivies dans le temps pour produire des traces d'ions correspondant aux peptides d'intérêt. Tous les ions produits MS2 dérivés de peptides prédéfinis (précurseurs MS1) sont enregistrés dans le temps pour générer les traces chromatographiques ioniques. (Panneau inférieur) La HRM et la SWATH MS sont deux méthodes d'acquisition indépendantes des données fondamentalement similaires qui utilisent respectivement un spectromètre de masse précis à haute résolution de type Orbitrap et un TOF quadripolaire. En mode DIA, les traces ioniques multiplexées sont acquises en effectuant des cycles répétés dans des fenêtres d'isolement de précurseurs consécutifs prédéfinis et en contrôlant tous les fragments d'ions.

Adapté de Caron et al (2015) [54].

### **1.3.1.3.1 - L'acquisition de données**

#### **1.3.1.3.1.1 - L'acquisition dépendante de données (DDA)**

Lancées par Donald Hunt au début des années 1990, les analyses des peptides associés au CMH par SM à analyse dépendante des données ont permis d'acquérir des connaissances de bases sur les motifs de liaison des peptides des molécules du CMH[44]. Le principal avantage de cette méthode d'analyse est qu'elle rend possible l'identification de milliers de peptides en un court laps de temps[54, 107].

En mode DDA, les ligands ionisés des peptides du CMH sont d'abord détectés dans un balayage d'étude (balayage MS1). Les ions précurseurs les plus abondants détectés dans le balayage MS1 sont ensuite sélectionnés pour être fragmentés par dissociation induite par collision, par dissociation à haute énergie de type faisceau ou par transfert d'électrons. Les ions fragmentés résultants sont finalement détectés et enregistrés dans le balayage MS2 (**Fig. 7. DDA-MS**). Pour l'identification des séquences, les spectres MS2

enregistrés sont comparés à une base de données à l'aide de moteurs de recherche commerciaux ou de type "open source"[44, 54]. Un des désavantages de cette méthode est que dû à la nature courte des peptides associés aux CMH ainsi que la faible proportion des peptides contenant des acides aminés basiques, de nombreux peptides de classe I sont inefficacement ionisés et les méthodes de fragmentation standards génèrent des spectres MS2 peu prévisibles et pauvres en informations. Ainsi, environ 10% des spectres MS2 acquis contiennent des informations suffisantes pour l'attribution correcte de la séquence peptidique. Toutefois, certaines avancées dans les méthodes de fragmentation, notamment celles de type "dissociation par transfert d'électrons et collision à haute énergie" ont conduit à des assignations de peptides HLA de classe I de haute confiance pour 39% des spectres MS2 acquis[54, 108]. Bien que l'immunopeptidomique basée sur la SM ait grandement bénéficié des progrès réalisés ces dernières années en matière d'instrumentation et d'optimisation de protocoles, il est désormais évident que les spectromètres de masse fonctionnant en mode DDA sont moins bien adaptés à la résolution de problèmes nécessitant la comparaison d'ensembles de données reproductibles pour de nombreux échantillons ou conditions. Trop peu de peptides composant le lysat cellulaire sont sélectionnés pour être fragmentés, ce qui résulte en une identification non-reproductible des peptides lorsque le même échantillon est analysé à plusieurs reprises. Une proportion significative des peptides identifiés peut varier lorsqu'un même échantillon est analysé à plusieurs reprises[54]. Les stratégies de SM ciblées peuvent pallier les lacunes que cette technique apporte[54, 109, 110].

#### **1.3.1.3.1.2 L'acquisition de données ciblées (*Targeted MS*)**

L'acquisition de données ciblées se fait avec l'analyse d'ensemble prédéfinis de peptides, et ce, avec un niveau élevé de sensibilité et de reproductibilité. Les méthodes ciblées sont principalement réalisées en mode S/MRM (*Selective/Multiple Reaction Monitoring*) sur un spectromètre de masse Orbitrap quadripolaire à faible résolution ou en mode PRM (*Parallel Reaction Monitoring*) sur un spectromètre de masse Orbitrap quadripolaire à haute résolution[54, 111, 112]. Pour une analyse ciblée, des peptides

spécifiques et des fragments de peptides dans une liste prédéfinie sont sélectionnés sur la base de leurs valeurs  $m/z$  (masse/charge) prédéterminées par les premier et troisième quadrupoles d'un instrument quadrupolaire triple (**Fig. 7. Targeted-MS**). Ceci permet d'enregistrer l'intensité de chaque transition durant la chromatographie. Une fois que les chromatogrammes ont été acquis, des logiciels de traitement automatique des données SRM peuvent être appliqués pour calculer la probabilité que le peptide ciblé ait été détecté dans l'échantillon et déterminer sa quantité. Cette méthode est reconnue comme étant l'étalon d'or au niveau de la quantification d'ensembles de données prédéfinies[54, 113].

Malgré ses avantages, principalement au niveau quantitatif, la principale restriction de cette technique est qu'elle est limitée par sa capacité à détecter seulement quelques centaines de peptides par injection d'échantillon et n'est donc pas idéalement adaptée à la quantification exhaustive des immunopeptidomes possédant plusieurs dizaines, voire centaines de milliers de peptides.

Pour surmonter cette limitation, de nouvelles méthodes de spectrométrie de masse, connues sous le nom de DIA, ont récemment vu le jour et devraient accélérer la compréhension des immunopeptidomes.

#### **1.3.1.3.1.3 - L'acquisition de données indépendante de données (DIA).**

Parmi les différentes méthodes d'acquisition indépendantes de données, le SWATH-MS (de l'anglais Sequential Window Acquisition of all Theoretical Mass Spectra) a été revu par Sajic, T. et al (2015) et a récemment été appliquée à l'analyse d'immunopeptidomes provenant d'une série de sources biologiques[54, 110, 113].

Brièvement, ce mode d'acquisition convertit les peptides de l'échantillon en une carte virtuelle composée de spectres MS2 multiplexés dérivés de la fragmentation de tous les ions précurseurs. La collecte des données est donc complète et les informations quantitatives peuvent être analysées ultérieurement. Contrairement aux analyses en mode DDA, chaque spectre MS2 enregistré n'est pas représenté par un seul peptide, mais par un enregistrement multiplexé des fragments d'ions dérivés de tous les peptides éluant

en temps réel dans la plage  $m/z$  prédéfinie par la fenêtre de précurseurs (**Fig. 7. DIA-MS**). Ce séquençage continu de tous les peptides détectables entrant dans le spectromètre de masse permet de collecter des données avec moins de valeurs manquantes qu'en utilisant le mode DDA conventionnel et des niveaux de multiplexage beaucoup plus élevés que ceux offerts par le S/MRM [54, 113].

Les grands avantages de cette méthode sont que le SWATH-MS en mode DIA permet la mesure reproductible et quantitativement précise des protéines dans plusieurs échantillons. Essentiellement, cette méthode possède tous les avantages des méthodes S/MRM en plus de permettre l'analyse rapide de centaines de milliers de peptides à travers plusieurs échantillons[54, 94, 114].

Grâce à la méthode du SWATH-MS, Caron et al. (2015) ont démontré, pour la première fois, la faisabilité d'un effort conjoint au sein de plusieurs laboratoires pour construire des bibliothèques spectrales reproductibles des allèles HLA à travers plusieurs échantillons biologiques[110]. Ils ont d'abord créé un flux de travail informatique pour soutenir l'identification, la classification/annotation, la visualisation et le stockage des données peptidomiques HLA d'une manière dépendante des allèles. Permettant (1) l'annotation systématique des peptides à leur allèle HLA respectif, (2) la visualisation d'ensembles de données peptidomiques HLA et (3) la génération de bibliothèques spectrales de peptides spécifiques de l'allèle HLA de classe I, qui peuvent être converties en "*assay libraries*" de haute qualité pour le traitement des données SWATH, une approche émergente pour des mesures quantitatives reproductibles, cohérentes et précises des peptides. Ainsi, le travail conjoint de ce consortium ainsi que les ressources informatiques de données présentées dans cette étude constituent un premier pas vers des mesures hautement cohérentes, reproductibles et quantitativement précises des immunopeptidomes à travers multiples échantillons.

### 1.3.1.3.2 - Analyses de données

Différentes techniques ont été décrites pour identifier des protéines à partir de données de spectrométrie de masse. Les approches les plus courantes s'appuient sur les bases de données disponibles pour y faire correspondre les données expérimentales.

Un des problèmes inhérents à l'utilisation de cette technique est qu'une grande proportion des spectres identifiés sont des contaminants. Leur analyse fait perdre d'énormes quantités de temps de calcul[115]. Les méthodes en protéomique ont été adaptées à l'étude de l'immunopeptidome mais elles s'accompagnent d'un certain nombre d'inconvénients. Premièrement, les banques de données deviennent de plus en plus volumineuses, augmentant ainsi le temps nécessaire à traiter un échantillon. De plus, dû à l'absence de trypsinisation des protéines, le temps requis pour analyser un échantillon croît de manière exponentielle. En raison de ces deux problèmes, le temps nécessaire pour effectuer une analyse peut augmenter considérablement comparativement à un échantillon protéique standard. Il peut augmenter davantage si des modifications post-traductionnelles doivent être prises en compte dans la recherche[60]. Par conséquent, un compromis est souvent nécessaire et le taux de fausses découvertes accepté au niveau des identifications peut atteindre 5 % au lieu du taux standard de 1 % utilisé dans les expériences standard de protéomique en spectrométrie de masse[60].

Avec la disponibilité croissante d'échantillons cliniques appropriés, il existe désormais une forte demande pour des protocoles automatisés et reproductibles permettant le traitement rapide de données expérimentales ainsi que l'analyse à grande échelle de données publiques, par exemple, à partir de banques de données en protéomique ou immunopeptidomique telles que PRIDE ou SystemMHC [54, 116]. De plus, le développement de bases de données sur les peptides du CMH, comme par exemple, l'atlas des ligands HLA (SystemMHC atlas), a grandement contribué à l'accélération du partage des données et reflète la nature collaborative du domaine [97, 98, 117]

#### **1.3.1.3.2.1 - Traitement des données brutes**

L'identification de séquences peptidiques à partir de données MS/MS brutes est lente et fastidieuse lorsqu'elle est effectuée manuellement. L'approche la plus logique et populaire consiste à comparer les données MS/MS expérimentales non interprétées à des bases de données. Les données acquises par les spectromètres de masse à haute résolution sont interprétées à l'aide d'algorithmes qui permettent d'assigner les spectres de masse aux séquences d'acides aminés. Une grande variété de logiciels pour l'interprétation des spectres de fragments sont disponibles, allant des logiciels commerciaux aux algorithmes "open-source" tels que MASCOT [118], SEQUEST[119], X! Tandem<sup>82</sup> et Maxquant<sup>83</sup>[57]. Plus récemment, des logiciels de séquençage de novo tels que PEAKS X, sont utilisés pour extraire des informations sur les séquences d'acides aminés sans l'utilisation de bases de données[120]. Ce logiciel utilise des informations préliminaires provenant de recherches de novo afin de limiter l'espace de recherche au niveau des bases de données, améliorant ainsi la vitesse et les performances des moteurs de recherche[57, 120].

#### **1.3.1.3.2.2 - Analyses de prédiction de liaison peptidique**

Parmi les logiciels les mieux documentés et les plus utilisés pour prédire la liaison des peptides au CMH de classe I figurent NetMHC et NetMHCpan. Ces outils ont suscité un intérêt au cours des dernières années en raison de l'intérêt récent pour l'identification des néoantigènes dans le domaine de l'immunothérapie personnalisée[47, 51]. Ces outils analysent les motifs peptidiques afin d'obtenir des prédictions de liaison utilisées pour annoter les peptides de manière suivante : pour la classe I, le rang des liants forts est de 0,5, celui des liants faibles est de 2,0 et celui des non liants est  $> 2,0$  ; pour la classe II, le rang des liants forts est de 2,0, celui des liants faibles est de 10 et celui des non liants est  $> 10$ . Ces logiciels, en exploitant à la fois les peptides et la séquence HLA primaire comme informations d'entrée pour des prédictions basées sur des réseaux neuronaux artificiels, ont démontrés avec succès leur potentiel pour prédire l'affinité d'interaction de n'importe quel peptide avec n'importe quelle molécule HLA[51].

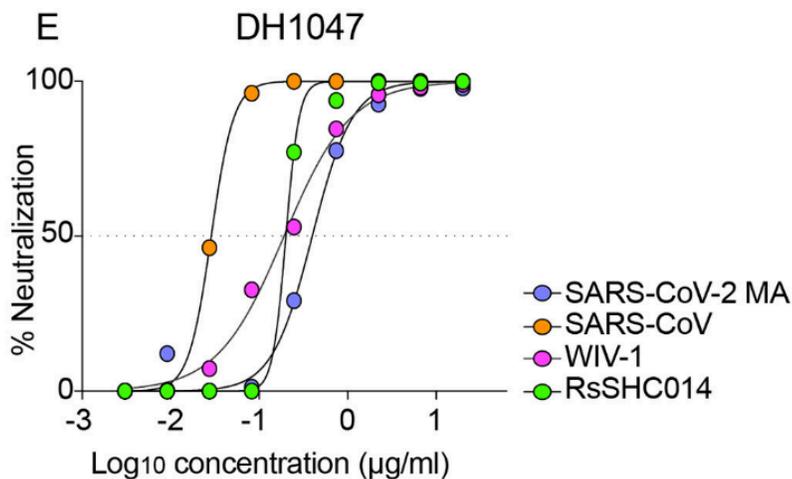
#### **1.3.1.3.2.3 - Analyse intégrée des peptides associés au CMH : MHCvizPipe (MVP)**

Un des problèmes majeurs dans l'analyse des peptides associés au CMH à partir de données brutes de spectrométrie de masse est que l'utilisation des logiciels requis peut rester relativement difficile pour les non-experts disposant d'une expérience et de ressources informatiques limitées. À ce problème s'ajoutent l'impossibilité d'analyser plusieurs échantillons de manière simultanée et par-dessus tout, un manque d'outil de contrôle de la qualité, de la spécificité des échantillons et de la reproductibilité des opérations à travers plusieurs laboratoires et instruments. À ces fins, un logiciel facile à installer et à utiliser pour l'évaluation rapide de la qualité et de la spécificité des ensembles de données immunopeptidomique a été développé par Kovalchik et al: MVP[117]. Ce logiciel libre et gratuit peut être utilisé par n'importe quel laboratoire pour le traitement rapide de l'information et visualisation de multiples ensembles de données immunopeptidomiques.

L'utilisation de ce logiciel permet d'identifier le nombre de peptides, la longueur des peptides, le score d'affinité de liaison HLA prédit en fonction du sous-type HLA de l'échantillon analysé et le motif de liaison du peptide. De telles informations permettent un contrôle de la qualité des échantillons et des manipulations en assurant la spécificité et la reproductibilité des résultats. MVP accélère considérablement le processus d'analyse pour évaluer la qualité des ensembles de données immunopeptidomiques provenant de n'importe quelle source biologique et permet une comparaison rapide des différents protocoles, facilitant ainsi leur optimisation[117]. Le domaine de l'immunopeptidomique présente un grand intérêt scientifique et clinique dans le traitement des cancers, maladies auto-immunes et infectieuses[60], mais souffre d'un manque de protocoles normalisés et de contrôle de la qualité et de la spécificité des échantillons. Des outils tels que MVP sont essentiels au développement de nouvelles méthodologies afin de libérer tout le potentiel de l'immunopeptidomique dans la recherche biomédicale[45, 49, 50, 54, 60, 96, 97, 109, 110, 113, 117].

#### 1.4 - Immunopeptidomique et pandémie de coronavirus

Au cours des trois dernières décennies, les  $\beta$ -Coronavirus ont été la cause des éclosions de SRAS-CoV-1 (*Severe Acute Respiratory Syndrome Coronavirus 1*), MERS (*Middle Eastern Respiratory Syndrome*) et de la pandémie mondiale de SRAS-CoV-2 qui a jusqu'à ce jour infecté 223 millions de personnes et causé la mort de 4.6 millions d'entre eux (<https://covid19.who.int/> - visité le 8 septembre 2021). Plusieurs autres membres de cette famille de virus qui circulent actuellement chez les civets, chauve-souris et pangolins sont similaires génétiquement et possèdent une homologie au niveau de l'utilisation du récepteur ACE2 dans le cycle d'entrée virale, leur donnant ainsi le potentiel d'être transmis à l'homme [121, 122]. La preuve de concept de la possibilité d'un vaccin neutralisant un large spectre des membres du sous-génome des *Sarbecovirus* a été faite lors de la découverte d'anticorps ayant un potentiel de neutralisation croisée envers plusieurs  $\beta$ -Coronavirus. L'une des cibles des anticorps croisés est le domaine de liaison aux récepteurs (RBD) de la protéine « spike » [92, 123, 124]. Par exemple, l'anticorps DH1047 est un anticorps croisé qui cible le RBD et neutralise de manière croisée le SRAS-CoV, le SRAS-CoV-2 et certains coronavirus de la chauve-souris (**Fig. 8**) [92, 125]. Plus récemment, des études ont identifiés des clones de cellules T à réaction croisée chez des individus non-exposés au COVID 19, ce qui étaye le phénomène d'empreinte antigénique par une exposition antérieure au Coronavirus courants qui circulent chez l'humain[124, 126].



**Figure 8 :** Neutralisation des virus SARS-CoV-2, SARS-CoV, WIV-1 et RsShC014 (souches circulantes chez les chauve-souris) par l'anticorps neutralisant DH1047. Adapté de Martinez, D et al (2021) [127]

La production d'anticorps neutralisant suite à l'inoculation par un vaccin est sans aucun doute la barrière la plus efficace pour empêcher la propagation d'un virus. Contrairement à la réponse des cellules T qui s'enclenche seulement une fois qu'un individu est infecté, la présence d'anticorps peut empêcher l'apparition de l'infection[92, 95]. Toutefois, de plus en plus d'éléments indiquent que la réponse des cellules T spécifiques du SRAS-CoV-2 jouent un rôle crucial dans le contrôle et dans la sévérité des symptômes associés à la maladie de COVID-19[128]. Les cas les plus légers de COVID-19 ont été associés à des réponses coordonnées des anticorps, des cellules T CD4+ et CD8+, tandis que les cas graves ont été corrélés à un manque de coordination des réponses cellulaires et des anticorps et à une cinétique retardée de la réponse adaptative[129]. L'identification des épitopes du SRAS-CoV-2 qui engagent une forte réponse T est donc un élément clé pour le design de vaccins de seconde génération [130].

Les résultats présentés ici soulignent l'importance d'étudier la spécificité antigénique pour faciliter notre compréhension des mécanismes qui soutiennent une immunité efficace ainsi que d'identifier les épitopes les plus immunogéniques communs chez les différents membres de la famille de Coronavirus [92, 126]. Weingarten-Gabbay et al. (2021)[62] ont rapporté le premier immunopeptidome HLA-I du SRAS-CoV-2 dans deux lignées cellulaires à en utilisant la spectrométrie de masse. De manière intéressante, ils ont trouvé des peptides dérivés non seulement des cadres de lecture ouverts canoniques, mais aussi des peptides provenant de séquences hors-cadre qui ne sont pas pris en compte par les vaccins actuels [62].

La plupart des études qui se sont intéressées à l'interaction entre les lymphocytes T et les antigènes du SRAS-CoV-2 ont utilisé jusqu'à présent des approches bio-informatiques pour prédire la liaison peptide-HLA-I[131]. Bien que la prédiction de la liaison HLA-I soit sans aucun doute un outil utile pour identifier des antigènes présents

dans la niche des molécules HLA, elle a des limites. Ces limites sont principalement dues au fait que la présentation de l'antigène se fait suite à un processus comprenant de nombreuses étapes au niveau du façonnage des peptides : de la dégradation par le protéasome au clivage des peptide par des aminopeptidases suivi de la translocation dans le réticulum endoplasmique. Également, ces modèles informatiques ne tiennent pas compte de la capacité des virus à influencer la nature des peptides présentés et à interférer avec la présentation de l'antigène. La valeur prédictive positive moyenne obtenue pour l'ensemble des allèles HLA par approche computationnelle se situe aux alentours de 64% [62, 132].

Pour soutenir l'importance de l'identification *in-vivo* des peptides viraux, Weingarten-Gabbay et al. (2021) ont utilisé HLATHENA afin de tenter de prédire de manière rétrospective les peptides de 8 à 11-mer qui composent les protéines de SRAS-CoV-2 contre le complément des allèles HLA-I exprimés par les deux lignées cellulaires utilisées dans leur étude (A549 et HEK293T). Sur les 36 peptides identifiés par SM, 23 avaient un rang percentile prédit inférieur à 0,5, et 31 avaient un rang percentile inférieur à 2. Parmi les 39 875 peptides SRAS-CoV-2 entre 8 et 11 résidus, 14 des 18 peptides HLA-I des cellules A549 et 11 des 18 peptides des cellules HEK293T avaient des scores de rang percentile parmi les 1 000 premiers peptides viraux (1,5 % et 1,7 % pour les cellules A549 et HEK293T, respectivement). Pour tenir compte de la variabilité des niveaux d'expression des différentes protéines virales, ils ont répété cette analyse pour chaque protéine source desquelles ils ont identifié des peptides. Ils ont constaté que 16 des 36 peptides ont obtenu un score parmi le top 10 parmi tous les peptides de 8 à 11 résidus de la protéine source, et 21 parmi les 20 premiers. Ces observations suggèrent que, bien qu'un schéma de prédiction d'épitope *in silico* qui nomme les 10-20 meilleurs peptides de chaque protéine virale récupère 50 % (16-21 sur 36) des épitopes observés avec une très haute priorité, cette liste ne comprendrait toujours que 5-10 % de vrais positifs LC-MS/MS[62]. L'immunopeptidomique basée sur la spectrométrie de masse est donc une méthode *in-vivo* de choix pour découvrir des peptides présentés de manière endogène qui pourrait échapper aux méthodes de prédictions bio-informatiques utilisés de manière courante.

Idéalement, ce travail aurait utilisé le même modèle cellulaire et SRAS-CoV-2 pour développer une plate-forme d'étude de l'immuno-peptidome de cellules infectées par SRAS-CoV-2, et par le fait même, évaluer la reproductibilité inter-établissements de cette analyse. Toutefois, des contraintes administratives nous ont empêchées de recevoir les formations nécessaires à temps. Toutefois, avec la collaboration du laboratoire du Dr Roger Lippé du centre de recherche CHU St-Justine, nous avons pu utiliser la lignée cellulaire MRC-5 et le coronavirus HCoV-OC43 comme modèle d'étude pour déterminer la faisabilité de l'analyse de l'immuno-peptidome de cellules infectées par un coronavirus. L'étude de ce virus est pertinente car il provient de la même famille que SRAS-CoV-2, il infecte également par voie aérienne et il possède une grande homologie avec lui[133]. Les cellules MRC-5 sont aussi un modèle d'étude pertinent car ce sont des cellules qui proviennent du poumon, organe principalement affecté par les coronavirus[134].

**Tableau 1 :** Épitopes du SRAS-CoV-2 avec les prédictions HL Athena correspondantes pour les 6 allèles HLA-I exprimés par les cellules A549 et les 3 allèles HLA-I dans les cellules HEK293T qui contrairement aux cellules A549, proviennent d'une lignée monozygote. Adapté de Weingarten-Gabbay et al (2021) [62]

**A**

	peptide	viral protein	best predicted rank	best predicted allele	rank within all viral peptides	%rank within all viral peptides	rank within protein	%rank within viral protein	# of alleles with rank <=2	all alleles with rank <=2
A549	VATSRTLSY	M	0.0007	C1601	2	0.005	1	0.1171	2	C1203,C1601
	FAVDAAKAY	nsp10	0.0027	C1601	8	0.020	1	0.1916	3	A2501,C1203,C1601
	FASEAARVV	nsp2	0.0049	C1601	18	0.045	2	0.0794	2	C1203,C1601
	DEFVVVTV	ORF9b	0.0083	B1801	24	0.060	2	0.5650	2	B1801,B4403
	NATNVVIVK	S	0.0180	C1203	45	0.113	5	0.0981	2	C1203,C1601
	EIKESVQTF	nsp2	0.0184	A2501	47	0.118	3	0.1191	1	A2501
	EEFEPSTQYEE	nsp3	0.0194	B1801	49	0.123	18	0.2324	3	A2501,B1801,B4403
	SEFSSLPSY	nsp8	0.0246	B1801	59	0.148	1	0.1319	2	B1801,B4403
	TVIEVQGY	nsp3	0.0528	A2501	128	0.321	34	0.4389	1	A2501
	LATNNLVVM	nsp2	0.0848	C1203	189	0.474	18	0.7149	2	C1203,C1601
	TTTIKPVTY	nsp3	0.0869	C1601	196	0.492	51	0.6584	4	A2501,B1801,C1203,C1601
	EILDITPCSF	S	0.0955	A2501	210	0.527	32	0.6279	2	A2501,B1801
	TAQNSVRVL	nsp2	0.1829	C1601	398	0.998	41	1.6283	2	C1203,C1601
	LEDKAFQL	ORF9b	0.2740	B1801	613	1.537	6	1.6949	2	B1801,B4403
	VGYLQPRTF	S	0.6251	C1601	1473	3.694	185	3.6303	1	C1601
	QLTPTWRVY	S	0.7696	C1601	1835	4.602	238	4.6703	3	A2501,C1203,C1601
KNIDGYFKIY	S	0.8731	B1801	2074	5.201	265	5.2002	4	B1801,B4403,C1203,C1601	
EILDITPCSEFG	S	4.5200	A2501	10633	26.666	1274	25.0000	0	unassigned	
HEK293T	YLFDESGEFKL	nsp3	0.0052	A0201	6	0.015	1	0.0129	2	A0201,C0702
	KRVDWTIEY	nsp14	0.0180	C0702	13	0.033	1	0.0482	1	C0702
	APHGHVMVEL	nsp1	0.1099	B0702	95	0.238	3	0.4373	1	B0702
	IRQEEVQEL	ORF7a	0.1322	C0702	122	0.306	3	0.6667	1	C0702
	YLNSTNVTI	nsp3	0.1354	A0201	126	0.316	24	0.3098	1	A0201
	SVVSKVVKV	nsp15	0.1482	A0201	145	0.364	5	0.3704	1	A0201
	GLITLSYHL	S.iORF1/2	0.2323	A0201	245	0.614	3	3.3333	1	A0201
	ELPDEFVVVTV	ORF9b	0.4190	A0201	489	1.226	9	2.5424	1	A0201
	SLEDKAFQL	ORF9b	0.4349	A0201	512	1.284	10	2.8249	1	A0201
	FGDDTVIEV	nsp3	0.5348	A0201	629	1.577	114	1.4717	1	A0201
	KAFQLTPIAV	ORF9b	0.5696	A0201	682	1.710	12	3.3898	1	A0201
	ELPDEFVVV	ORF9b	0.7673	A0201	1006	2.523	18	5.0847	1	A0201
	NLNESLIDL	S	0.8444	A0201	1131	2.836	140	2.7473	1	A0201
	STSAFVEIV	nsp2	1.3088	A0201	1866	4.680	84	3.3360	1	A0201
	GPMVLRGLIT	S.iORF1/2	2.1237	B0702	3184	7.985	18	20.0000	0	unassigned
	MLLGSMLYM	S.iORF1	2.1581	A0201	3238	8.120	23	18.8525	0	unassigned
	APRITFGGP	N	3.0563	B0702	4781	11.990	151	9.1961	0	unassigned
	AGDTTTTIV	nsp5	3.4456	A0201	5467	13.710	129	10.8403	0	unassigned

## 1.5 - Contexte de recherche

### 1.5.1 - Objectifs de recherche

Les molécules du CMH sont les pièces maîtresses de la présentation des antigènes, de l'immunité adaptative et ont un rôle clé dans le contrôle des maladies infectieuses. La qualité des données est la pierre angulaire de recherches exigeant répétabilité et reproductibilité. Dans ce travail, nous évaluons l'efficacité d'un protocole d'isolation des peptides associés aux molécules du CMH conjointement à l'utilisation d'une plate-forme de spectrométrie de masse dans le but de caractériser l'immunopeptidome de cellules infectées par un coronavirus. La connaissance de la composition des peptides viraux qui sont dégradés pour être présentés à la surface des molécules du CMH est fondamentale et revêt une importance clinique significative lorsqu'il s'agit de concevoir des vaccins ARNm de deuxième génération.

Ce travail a pour but d'établir une plate-forme d'étude de l'immunopeptidome de cellules infectées par des coronavirus. Étant donné qu'il a été démontré que les logiciels de prédiction *in-silico* ont des lacunes, l'identification *in-vivo* des peptides viraux présentés sur les molécules du CMH-I pourrait mener à l'amélioration de la sélection des séquences des épitopes viraux intégrés dans de futurs vaccins de deuxième génération contre les coronavirus. Ce travail comporte 5 objectifs adressés dans les chapitres 2 (objectifs 1 et 2), 3 (objectif 3) et 4 (objectif 4, 5).

1 - Déterminer l'abondance absolue des molécules du CMH-1 sur une batterie de lignées cellulaires utilisés de manière courante en immunologie.

2 – Isoler des peptides du CMH par purification d'immuno-affinité à partir de lignées cellulaires de références servant de contrôle de qualité.

3 – Utilisation du logiciel MHCvizPipe comme outil de contrôle de la qualité pour l'évaluation rapide de la spécificité et reproductibilité des immunopeptidomes isolés à l'objectif 2.

4 – Appliquer les méthodes et outils de contrôle de qualité des objectifs 1, 2 et 3 pour évaluer la reproductibilité de l'immunopectidome de cellules A549-ACE2/TMPRSS2 inter-établissement

5 – Appliquer les méthodes et outils de contrôle de qualité des objectifs 1, 2 et 3 pour analyser l'immunopectidome de cellules MRC-5 avant et après infection avec Coronavirus HCoV-OC43 qui circule chez l'humain.

## **Chapitre 2 – Isolation des peptides du CMH par purification d’immunoaffinité : contrôle de la qualité à partir de lignées cellulaires de références.**

### **2.1 - Lignées de références - contrôle de la qualité des immunopeptidome**

Avant de comparer l’immunopeptidome de cellules naïves à celui de cellules infectées par un virus, il est important de s’assurer de la qualité et reproductibilité de la démarche expérimentale utilisé pour isoler et analyser les peptides situés sur les molécules du CMH.

Pour se faire, nous avons utilisé des cellules couramment utilisées en immunologie afin de s’assurer de la qualité et reproductibilité des protocoles d’isolation des peptides par immunopurification.

#### **2.1.1 – Choix des lignées cellulaires appropriés**

Bien que certains groupes spécialisés en immunopeptidomique se penchent sur la caractérisation du polymorphisme HLA et la liaison de peptides spécifiques à certains allèles du CMH, peu d’information sont disponible sur la différence entre l’abondance des molécules du CMH à la surface de différentes lignées cellulaires.

Avant de procéder à l’isolation et à la caractérisation des peptides du CMH, nous avons entrepris de déterminer l’abondance absolue des molécules HLA sur plusieurs modèles cellulaires avec des techniques de cytométrie établies. Cette étape est cruciale au niveau du contrôle de la qualité car le niveau d’expression des molécules du CMH varie grandement entre les différents tissus et modèles cellulaires. Étant donné qu’une quantité se situant entre 6 et 10 mg de protéines est requise pour un procédé d’immunopurification affichant des résultats viables et reproductibles, le choix d’une lignée possédant une forte expression des molécules du CMH est critique à l’obtention de résultats significatifs[135]. Par la suite, nous avons tenté d’optimiser un protocole

d'isolation des peptides associés au CMH de classe I et II par purification par immunoaffinité à partir d'échantillons de contrôle de qualité. Alors que l'isolation des peptides associés à HLA-ABC à partir d'échantillons humains en utilisant l'anticorps W6/32 (BioXcell, InVivoMAb anti-human MHC Class 1 (HLA-A, HLA-B, HLA-C) #BE0079). Cet anticorps est bien établi et largement appliqué par de nombreux groupes de recherche par sa facilité d'utilisation et sa couverture complète des allèles A, B et C, par le même anticorps. L'isolation des peptides associés au CMH de classe I et II de la souris est relativement moins bien établi[57]. Par conséquent, des protocoles robustes pour l'isolation des ligands du CMH de souris sont nécessaires. Nous proposons ici un protocole optimisé pour l'isolement des peptides du CMH de classe I et des peptides du CMH de classe II à partir de deux lignées cellulaires de souris. Ainsi, la diffusion et l'application des protocoles proposés devraient faciliter la recherche immuno-peptidomique fondamentale et translationnelle dans divers modèles humains et murins.

### **2.1.2 – Lignées cellulaires modèles**

Nous avons effectué la quantification des molécules du CMH-1 sur les lignées cellulaires suivantes :

- JY, EL4 (murin), A20, THP-1, A549-ACE2, HEK-293T-ACE2 et Vero E6(Primate).

La quantification des divers types de cellules a été réalisé lors de plusieurs différents projets au sein de notre laboratoire, les cellules JY (cellules B lymphoblastoïdes humaines en suspension modifiées par virus EBV) et EL4 (cellules B, lymphome induit dans souris C57BL) sont considérés comme des lignées de référence et agissent en tant que contrôles positifs lors des expérimentations d'analyse du CMH-I sur des lignées cellulaires humaines ou de souris. Les Cellules A20 (cellule B, lymphome dérivé de souris BALB/c) jouent le même rôle de témoin positif pour les analyses du CMH-II. Les autres lignées mentionnées ont été analysés car elles étaient des modèles potentiels pour étudier l'infection de cellules par CoV-OC43 ou SRAS-CoV2, par leur susceptibilité au virus et par la pertinence du tissu infecté (cellules issues de tissu pulmonaire[135, 136]).

## 2.2 Matériel et méthodes.

### *Culture Cellulaire*

Préalablement à la culture cellulaire, le milieu de culture des cellules EL4 X-vivo 15 (numéro de catalogue : 04-744Q, Lonza) a été supplémenté par 25% de DMEM (numéro de catalogue : 319-005-CL, Multicell), 10% de FBS (numéro de catalogue : 98150, Multicell) et 1% de pen/strep (numéro de catalogue : 450- 201-EL, Multicell). Le milieu de culture des cellules JY, DMEM, a été supplémenté avec 10% de FBS et 1% de pen/strep. Le milieu de culture cellulaire A20 RPMI 1640 1X (numéro de catalogue : 350-000-CL, Multicell) a été supplémenté avec 10% de FBS, 1% de pen/strep, 1% de pyruvate de sodium 100mM (100X) (N. de catalogue 600-110-EL, Multicell), 1% de solution Hepes 1M (numéro de catalogue : 330-050-EL), 1% de D-glucose (code : 0188-500G, Amresco) et 0,1% de Beta-mercaptoethanol (code : 190242, MP Biomedicals).

Le premier jour de la culture cellulaire, la lignée de cellules lymphoblastoïdes de souris EL4 a étéensemencée dans deux flacons de 250 ml (Canted Neck Blue Vented Cap Tissue Culture Treated, numéro de référence : 353136, Falcon) avec  $10 \times 10^6$  cellules dans un volume de 25 ml de milieu X-vivo 15 ( $0,4 \times 10^6$  cellules/ml). Lors de l'utilisation de la lignée de cellules B lymphoblastoïdes humaines JY, deux flacons de 250 ml ont étéensemencés avec  $18,75 \times 10^6$  cellules dans un volume de 25 ml de milieu DMEM ( $0,75 \times 10^6$  cellules/ml). La lignée lymphoblastoïde de cellules B A20 a étéensemencée dans deux flacons de 250 ml avec  $7,5 \times 10^6$  cellules dans un volume de 25 ml de milieu RPMI 1640 1X ( $0,3 \times 10^6$  cellules/ml). Toutes les lignées cellulaires ont été incubées à 37°C avec 5% de CO<sub>2</sub> pendant une période de 48 heures, afin qu'elles atteignent une concentration d'environ  $2 \times 10^6$  cellules/ml). La densité cellulaire a été mesurée par un compteur de cellules automatisé Invitrogen countess II (numéro de catalogue : AMQAX1000) en utilisant 10 µl de milieu de culture mélangé à 10 µl de bleu trypan à 0,4% (Ref : EB7-001, NanoEntek). La concentration finale de cellules a été mesurée en utilisant la moyenne des deux chambres de la lame cellulaire. Les lignées cellulaires en suspension (JY, EL4, A20) ont été

récoltées en suivant le même protocole :  $100 \times 10^6$  cellules ont été récoltées et centrifugées à 1500 RPM pendant une période de 5 minutes à température ambiante dans des tubes à centrifuger coniques stériles en polypropylène de 50 ml (Ref : 352070, Falcon). Le milieu de culture a été retiré par aspiration et les culots cellulaires ont été lavés doucement (sans vortex) avec 5 ml de PBS (référence catalogue:311-010-CL). Après un second tour de centrifugation aux mêmes paramètres, le PBS a été retiré et les culots cellulaires ont été transférés dans des tubes de centrifugation stériles coniques de 15 ml en polypropylène haute clarté (Ref : 352096, Falcon) avec 3ml de PBS. Les tubes sont centrifugés à 1500 RPM pendant une période de 5 minutes. Après avoir éliminé le PBS par aspiration, les culots cellulaires ont été stockés à -80C. Les lignées cellulaires adhérentes (THP-1, A549-ACE2, HEK-293T, Vero E6), ont été décollées par trysinisation (N. de catalogue R001100, thermofisher) préalablement à la récolte.

Une liste de matériel ainsi qu'un horaire de culture optimisé se trouvent en annexe 1.

*Titrage d'anticorps conjugués à un fluorochrome pour le marquage de récepteurs de surface cellulaire sur diverses lignées cellulaires.*

Préalablement à l'analyse quantitative des récepteurs par immunofluorescence, un essai de titrage d'anticorps doit être effectué pour déterminer la quantité et la concentration d'anticorps donnant le signal le plus élevé de la population positive[137]. Le titrage des anticorps a été réalisé à l'aide d'un FACS CANTO II. Six tubes sont utilisés pour effectuer une dilution en série 1 :10 des anticorps primaires dirigés vers une molécule du CMH de cellules humaines ou de souris (EL4 :H-2 Db : Biolegend, Purified anti-mouse H-2 Db, clone H95, Cat:111502, H-2 Kb : Biolegend, Purified anti-mouse H-2 Kb, clone AF6-88.5, Cat:116502. JY: W6/32 : BioXcell, InVivoMAb anti-human MHC Class 1 (HLA-A, HLA-B, HLA-C)) dans un volume de 100  $\mu$ l de Tampon FACS (PBS-FBS1%). 6 culots de  $1 \times 10^6$  de cellules en suspension sont préparés. Les cellules sont incubées 15 minutes à température ambiante avec la solution d'anticorps primaire. Les cellules incubées sont

lavées et centrifugés deux fois (1500 RPM, 5 minutes) avec 3ml de tampon FACS à 4°C avant l'ajout de l'anticorps secondaire FITC (Dako, Polyclonal RabbitAnti-Mouse Immunoglobulins/FITC Code F0232) pendant 15 minutes, à température pièce et à l'abri de la lumière. Les cellules sont lavées et centrifugés deux fois à nouveau dans le tampon FACS avant d'être suspendu dans un volume de 250 µL de tampon FACS. Les 6 tubes de cellules marquées sont ensuite analysés dans un cytomètre FACS CANTO II. Les valeurs médianes de fluorescence (*Mean Fluorescence Index*) sont identifiées pour les 6 échantillons. Un graphique dans lequel les valeurs médianes de fluorescence des échantillons cellulaires sont représentées en fonction de la concentration d'anticorps de la dilution en série doit ensuite être créé. Il est ensuite possible de déterminer la concentration d'anticorps pour laquelle le rapport MFI/concentration d'anticorps est optimale en utilisant le point placé le plus près du début du plateau de la courbe.

#### *Analyse quantitative du MHC-1/2 par immunofluorescence sur diverses lignées cellulaires*

La coloration des cellules et les analyses par cytométrie de flux ont été réalisées à l'aide d'un FACS CANTO II avec l'utilisation de billes standard QIFIKIT (Dako Japan, Kyoto, Japon).  $1 \times 10^6$  cellules ont été incubées dans 100µl de milieu tampon FACS (PBS/FBS 1%) avec les anticorps appropriés (EL4 :H-2 Db : Biolegend, Purified anti-mouse H-2 Db, clone H95, Cat:111502, H-2 Kb : Biolegend, Purified anti-mouse H-2 Kb, clone AF6-88.5, Cat:116502. JY : W6/32 : BioXcell, InVivoMAb anti-human MHC Class 1 (HLA-A, HLA-B, HLA-C)) en conditions saturantes pendant 15 minutes à température ambiante. Un contrôle iso-type a été utilisé pour les lignées de cellules murines : Biolegend, Purified mouse IgG2b, k isotype ctrl, Cat : 402202 et un échantillon témoin non-coloré pour les lignées cellulaires humaine ont été utilisés pour établir toute fluorescence de fond. Les cellules incubées sont lavées et centrifugées deux fois (1500 RPM, 5 minutes) avec 3ml de tampon FACS à 4°C avant l'ajout de l'anticorps secondaire FITC pendant 15 minutes sans exposition à la lumière. Suite à la deuxième incubation, les cellules sont lavées et centrifugées deux fois à nouveau dans le tampon FACS. La quantification des protéines

MHC-1 de surface cellulaire a été rendue possible par l'utilisation des billes standard fournies dans le QIFIKIT qui sont traitées en parallèle avec les échantillons de cellules. L'intensité de fluorescence moyenne (*Mean Florescence Index*) est utilisée pour calculer une courbe d'étalonnage par rapport à la capacité de liaison des anticorps (*Antibody Binding Capacity*). Le détecteur laser de fluorescence verte a été réglé à 350 nm dans le FACSCantoll. La capacité spécifique de liaison aux anticorps (*Specific Antibody Binding Capacity*), correspondant au nombre moyen de récepteurs MHC de surface par cellule, a été mesurée en soustrayant la valeur ABC du contrôle iso-type ou témoins non-colorés aux échantillons cellulaires analysés.

Un protocole détaillé de cette expérimentation est inclus dans l'annexe 2.

#### *Isolation des peptides du CMH par purification d'immuno-affinité*

La génération d'échantillons de haute qualité, spécifique et dépourvus de contaminants est essentielle à la réussite de tout projet d'immunopeptidomique basé sur la spectrométrie de masse. Le protocole que nous avons décrit dans ce rapport a été adapté à partir de protocoles établis [57]. Nous avons notamment optimisé le protocole d'isolation des peptides de souris associés aux allèles H2Db/Kb. Si nécessaire, des étapes de fractionnement, de lyophilisation et/ou d'ultrafiltration à l'aide de filtres de 10-30 kDa, telles que décrites par d'autres auteurs, peuvent être utilisées pour augmenter le rendement des échantillons. À noter que le sous-type du CMH de classe I et II doit être connu pour isoler et identifier avec succès les ligands de classe I et II, ce qui peut être fait par un génotypage HLA à haute résolution [138]. Pour l'isolation des peptides du CMH-I sur des cellules humaines, l'anticorps W6/32 a été utilisé (W6/32 : BioXcell, InVivoMAb anti-human MHC Class 1 (HLA-A, HLA-B, HLA-C) #BE0079).

L'anticorps est couplé à des billes de sépharose activées par bromure de cyanogène. Pour se faire, les billes sont préalablement activées avec une solution de HCl 1mM pendant 30 minutes à température pièce. Après 30 minutes, les billes sont

centrifugées à 200 x g pendant 2 min afin de retirer le surnageant. 2mg d'anticorps sont ensuite joint aux billes dans un volume de 1 mL d'une solution de couplage d'anticorps (0.5 M NaCl, 0.1 M NaHCO<sub>3</sub>, pH 8.3) pour une durée de 120 minutes à RT (*Room Temperature*). Après deux heures, les perles sont centrifugées à 200 x g pendant 2 min avant d'en retirer le surnageant. Les sites d'ancrages non-utilisés par les anticorps sont ensuite bloqués en incubant les billes couplées aux anticorps dans 1 mL d'une solution de blocage (0.2 M glycine) pendant 60 minutes. Suite à cette période d'incubation, les billes sont lavées deux fois dans une solution PBS standard. La lyse des cellules est ensuite effectuée, un nombre de 100x10<sup>6</sup> cellules est recommandé pour chaque échantillon de cette expérience. 500ul de PBS est ajouté aux échantillons cellulaires. Par la suite, le culot cellulaire est incubé 60 minutes dans un volume égal à l'échantillon cellulaire d'une solution de PBS-Chaps 0,5% contenant des inhibiteurs de protéase. Les cellules sont centrifugées à 18000 x g pendant 20 min à 4°C. Le surnageant de lysat cellulaire est transféré avec les anticorps couplé aux billes, puis incubé 14-18 heures avec une rotation lente à 4°C. Les billes sont ensuite lavées tour à tour avec des solutions de : 150 mM NaCl, 20 mM Tris, pH 8.0, 400 mM NaCl, 20 mM Tris, pH 8.0 et Tris 20 mM pH 8.0 dans des collones Bio-Rad (Poly prep chromatography columns cat 731-1550). Les peptides liés au CMH sont ensuite élués en ajoutant 300 ul de TFA à 1%. Les étapes de déssalage et d'éluéion des peptides du CMH sont réalisées à l'aide d'une colonne C18 (The nest group Catalog # SEMSS18V). Une ronde de centrifugation de 1546 x g pendant 3min est effectué après l'ajout de chaque solution des étapes suivantes. La colonne est lavée avec 200uL de méthanol, 200uL de 80%ACN/0.1%TFA, 200uL of 0.1%TFA avant d'y ajouter la solution contenant les peptides élués. Une fois les peptides dans la colonne, laver l'échantillon avec 200uL de TFA à 0.1%. Finalement, éluer l'échantillon avec 200uL de 28%ACN/0.1%TFA trois fois pour un volume total de 600ul. L'échantillon est évaporé jusqu'à siccité en utilisant un speed-vac et il est gardé congelé jusqu'à ce que les échantillons soient analysés par LC-MS/MS (*Liquid Chromatography Tandem Mass Spectrometry*)

Un protocole détaillé de cette expérimentation est inclus dans l'annexe 3.

## *Identification des peptides du CMH de classe I et II par LC-MS/MS*

Les spectromètres de masse Orbitrap à haute performance et les spectromètres de masse quadripolaires à haute résolution sont les spectromètres de masse les plus appropriés pour l'analyse de l'immunopeptidome du CMH de classe I et II car ils affichent des résultats précis et reproductibles. Divers instruments de spectrométrie de masse en tandem existants fonctionnent selon des normes opérationnelles différentes et peuvent donner des résultats différents.

Les échantillons séchés sont solubilisés dans 50  $\mu$ L d'acide formique à 4%. Trois injections de 16  $\mu$ L pour chaque échantillon sont chargées et séparées sur une colonne en phase inverse faite maison (150- $\mu$ m i.d. par 250 mm de longueur, Jupiter 3  $\mu$ m C18 300 Å) avec un gradient de 5 à 30 % ACN-0,1 % FA et un débit de 600 nL/min sur un Easy nLC-1200 connecté à un Orbitrap Eclipse (Thermo Fisher Scientific, San Jose, CA). Chaque spectre MS complet est acquis à une résolution de, 120000, un AGC (*Automatic Gain Control*) de  $4 \times 10^5$  avec un mode automatique pour le temps d'injection. Il est suivi d'un tandem-MS (MS-MS) sur les ions précurseurs les plus abondants pendant une durée de 3s maximum. Les expériences de tandem-MS sont réalisées en utilisant une dissociation par collision à haute énergie de 30 %, une résolution de 30 000, un AGC de  $1,5 \times 10^5$ , et un temps d'injection de 300 ms. Les fichiers de données, provenant des trois injections par échantillons, sont traités à l'aide du logiciel PEAKS X (Peaks Pro V10.6, Bioinformatics Solutions, Waterloo, ON) en utilisant des bases de données de souris et d'humain (UniProtKB/Swiss-Prot (2019\_09)). Le paramètre "digestion enzymatique non spécifiée" est sélectionnée pour le paramètre enzymatique et les tolérances de masse sur les ions précurseurs et les ions fragments sont respectivement de 10 ppm et 0,01 Da. Les modifications variables sont la déamidation (NQ) et l'oxydation (M). Tous les autres paramètres de recherche sont les valeurs par défaut. Les listes finales de peptides sont filtrées en utilisant un ALC (*Alpha Cleavage*) de 80 % et un taux de fausse découverte de 1 % en utilisant le logiciel PEAKS.

## **2.3 - Résultats**

### **2.3.1 - Analyse quantitative du MHC-1/2 de surface par immunofluorescence**

La cytométrie en flux est un outil puissant pour détecter la présence de récepteurs ou de biomarqueurs sur des millions de cellules en une seule expérimentation. Des anticorps ayant une forte affinité et une grande spécificité pour leurs cibles sont marqués par fluorescence afin de mesurer l'expression des protéines ou biomarqueurs d'intérêt.

En raison de la variabilité des instruments, les résultats provenant d'analyses de cytométrie de flux traditionnel ont des limitations quand il est question de comparer des résultats inter-établissements. Avec l'aide du QIFIKIT, qui repose sur l'utilisation de populations témoins de billes avec un nombre établi de récepteurs, il est possible de générer une courbe d'étalonnage qui traduit le MFI d'un anticorps secondaire en nombre approximatif de molécules par cellule. Pour maximiser la précision de nos analyses, nous avons effectué un titrage d'anticorps sur les anticorps primaires destinés à être utilisés dans cette expérience. Les résultats représentent la moyenne de 3 expériences indépendantes.

En travaillant en condition de saturation, nous avons quantifié l'expression du CMH de Classe 1 avec l'anticorps W6/32 sur des cellules humaines JY et THP-1, nous avons aussi étudié l'expression du CMH-1 sur des cellules de souris EL4 avec les anticorps H-2 Db (Biolegend, Purified anti-mouse H-2 Db, clone H95, Cat:111502) et H-2 Kb (Biolegend, Purified anti-mouse H-2 Kb, clone AF6-88.5, Cat:116502). Les analyses du CMH de classe II sur les cellules JY ont été faites avec l'anticorps HLA-DR (BioXcell, InVivoMAb anti-human/monkey MHC class II (HLA-DR) #BE0306). Comme démontré à la figure 9, le nombre de molécules HLA-A/B/C par cellule JY et THP-1 est de 723611 et 284446 respectivement. L'expression du nombre de molécules du CMH-II sur les cellules JY de 478630 et 189055 pour les cellules A20. Le nombre des récepteurs des allèles H-2 Db et Kb sur les cellules murines est de 165518 et 355184 respectivement.

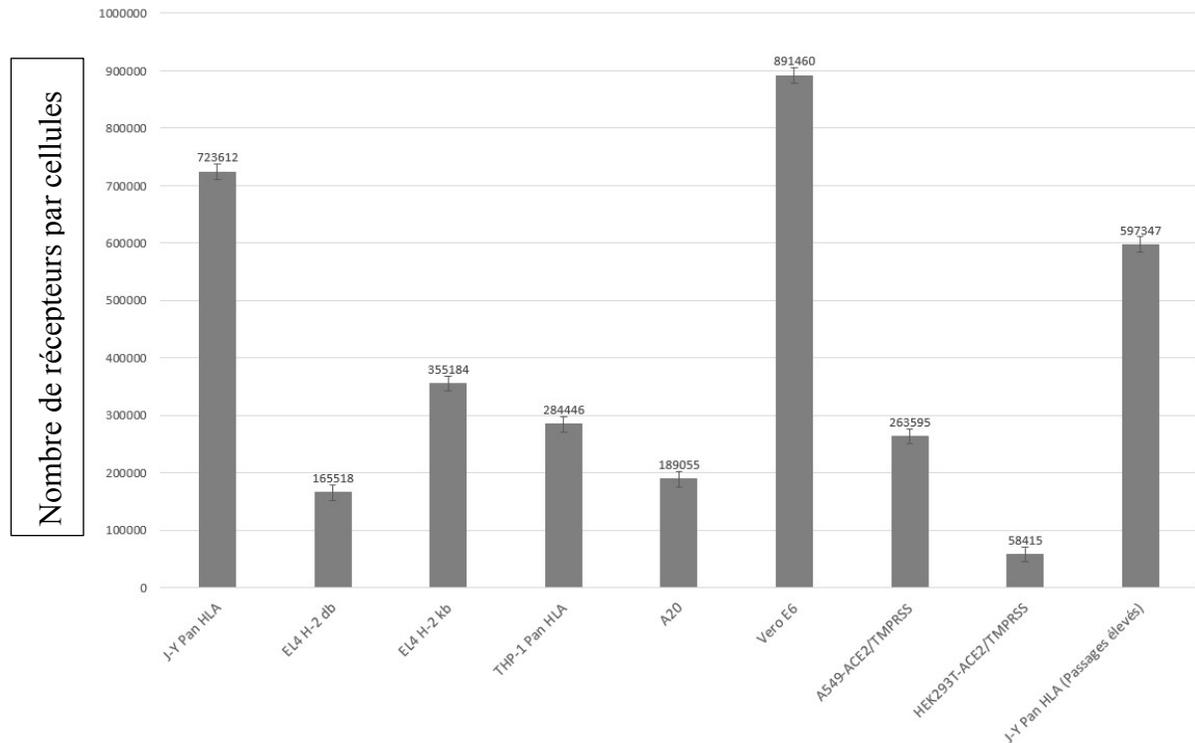
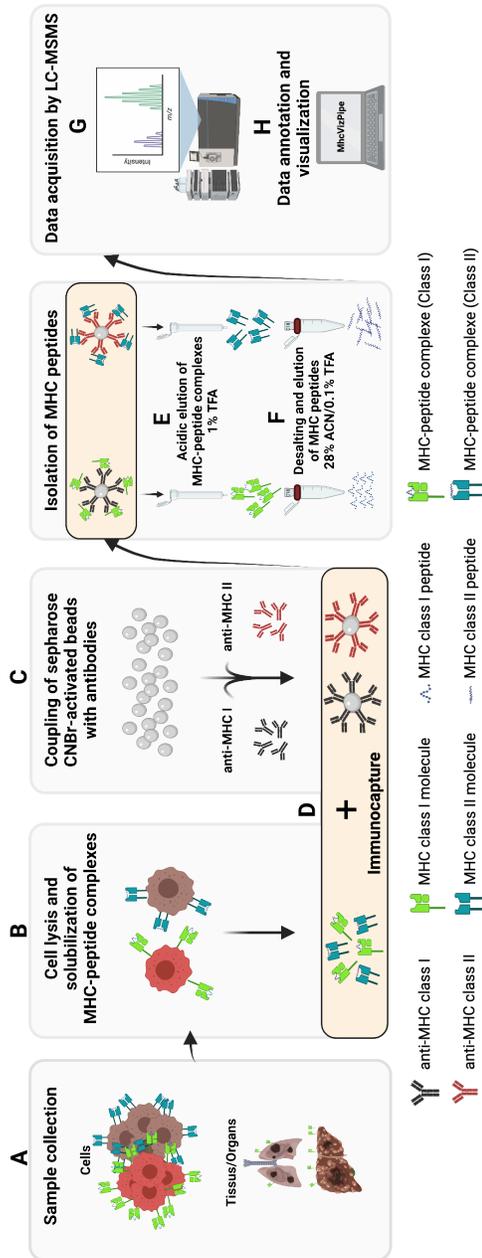


Figure 9 : Abondance de molécules du CMH I et II à la surface de diverses lignées cellulaires observée avec l'utilisation du QFIKIT.

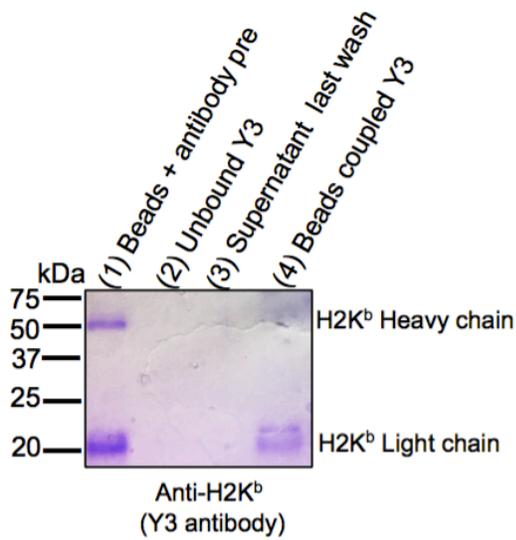
### 2.3.2 - L'isolation des peptides du CMH par purification d'immuno-affinité

Les étapes à suivre pour isoler les complexes CMH-peptide pour l'analyse d'immunopeptidomes par spectrométrie de masse sont illustrés dans la figure 10. Les prélèvements optionnels recueillis lors des différentes étapes de l'isolation des peptides par immunoaffinité (Annexe 3) ont été chargés sur des gels SDS-PAGE à 12 % et ils démontrent l'efficacité de la liaison entre l'anticorps Y3, qui reconnaît les molécules du CMH-1 de l'allèle H-2 Kb chez la souris, et les billes activées aux CNBr (**Fig. 11**) ainsi que la confirmation de l'isolation des complexes MHC-peptides suite à l'éluion à l'acide TFA (**Fig. 12**). En figure 11, l'efficacité de la liaison bille-anticorps est illustrée par une diminution significative du signal de l'intensité des chaînes légères et lourdes de H2-Kb lorsque les billes sont liées de manière covalente aux billes de CNBr (puits 4), par rapport à l'anticorps avant couplage (puits 1). En figure 12, Le signal de détection des complexes CMH-peptide en (C) en utilisant l'anticorps anti-chaîne lourde HLA-ABC anticorps (Abcam,

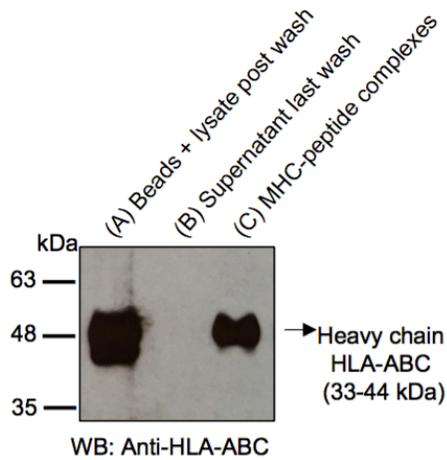
#ab 70328, 1:5000) a confirmé l'isolation des complexes CMH-peptide après l'élu-tion acide.



**Figure 10** : Schéma de la procédure complète pour l'isolation des peptides du CMH de classe I et II. Adapté de Sirois, I, Despault, J et al [135]



**Figure 11** : Coloration d'un gel d'agarose par bleu de Coomassie pour suivre l'efficacité de la liaison des anticorps Y-3 aux billes de sépharose activées par CNBr. Adapté de Sirois, I, Despault, J et al (2021)[135]



**Figure 12** : Immuno-transfert pour le suivi des complexes CMH-peptides extraits de cellules JY après élution acide des billes activées par CNBr couplées à des anticorps W6/32. Adapté de Sirois, I, Despault, J et al (2021)[135]

## 2.4 - Discussion

### 2.4.1 - Analyses de l'abondance des récepteurs de surface

Comme des études précédentes ont démontrés que des leucocytes bénins humains présentent en moyenne entre 75000 et 240000 récepteurs par cellules au niveau du CMH-I et 0 à 3200 au niveau du CMH-II, les résultats obtenus dans cette expérience nous permettent d'estimer que les cellules JY(723612 récepteurs), THP-1(284446 récepteurs) et EL4(520702 récepteurs) expriment des taux élevés de CMH-I et que les cellules A20 (189055 récepteurs) expriment de fort taux de CMH-II comparativement à des cellules bénignes (**Fig.9**)[139]. Nous avons trouvé une expression constante entre les répétitions lorsque qu'elles étaient effectuées en un court laps de temps, toutefois, lorsque que les cellules à l'étude ont été analysées subséquemment en tant que contrôle, les analyses ont montrés un plus grand taux de variation lors de l'utilisation de cellules récemment sortie de la cryo-congélation ou à passages plus élevés (**Fig.9**). Il serait donc recommandé de toujours utiliser des cellules acclimatées, sans toutefois qu'elles aient subit un nombre trop élevé de passages en culture cellulaire. Les protéines de surface cellulaire sont des régulateurs clés des processus cellulaires fondamentaux et, par conséquent, sont souvent à l'origine, ou profondément affectés de certaines conditions et pathologies. Malgré que NLRC5 et CIITA ont été identifiés comme régulateurs de l'expression des molécules du CMH I et II respectivement, les mécanismes pathologiques qui régulent la transcription des molécules du CMH et qui atténuent l'expression des gènes par modulation épigénétiques restent largement inconnus[140, 141]. Nous estimons que les procédures de quantification des molécules du CMH seront mis en évidence dans un avenir rapproché, non seulement afin de déterminer la propension de certaines pathologies à inhiber ou activer l'expression des gènes responsables de la transcription des molécules du CMH ou de sa sécrétion à la surface de la cellule, mais aussi comme potentiel outil de diagnostic qui permettrait d'identifier des anomalies dans la présentation du CMH associés à certaines maladies ou infections.

#### 2.4.2 - Optimisation des protocoles d'isolation des peptides du CMH

Le protocole actuel ne fournit pas de détails sur les étapes supplémentaires permettant de purifier davantage les peptides à la fin de la procédure d'isolement, par exemple le fractionnement des peptides par différentes méthodes ou l'ultrafiltration à l'aide de filtres de 10-30 kDa. Cependant, ces étapes pourraient être envisagées pour modifier et optimiser davantage la procédure d'isolation des peptides. Le type de billes et le type d'agent tampon d'élution acide utilisés pour éluer les complexes du CMH des billes peuvent également être modifiés pour améliorer les performances d'un anticorps dont les performances sont médiocres[101]. Les billes de sépharose activées par le CNBr constituent généralement une option viable car elles sont relativement peu coûteuses et font preuve de souplesse en termes de liaison avec divers types d'anticorps. Dans le protocole décrit dans cet ouvrage, les billes activées au CNBr se sont révélées relativement performantes avec cinq anticorps différents disponibles commercialement (M1, Y3, W6/32, L243 et M5). Les billes de sépharose (*4 Fast Flow*) à protéine A ou G ou sont également faciles à manipuler et, bien que relativement plus coûteuses, peuvent générer des résultats similaires. Un facteur à prendre en compte est l'affinité de l'anticorps pour la protéine A ou G. Également, les billes magnétiques de sépharose, quoique relativement coûteuses, sont également très faciles à utiliser[135]. Indépendamment du type de billes choisi, il est recommandé de prélever des aliquots des étapes optionnelles du protocole (annexe 3) et de réaliser des gels SDS-PAGE colorés au bleu de Coomassie pour suivre l'efficacité de la liaison de l'anticorps aux billes, comme le montre la figure 11. Le protocole peut être adapté et optimisé pour effectuer un criblage immunopeptidomique dans diverses matrices biologiques allant du plasma aux tissus frais ou congelés. En outre, la standardisation d'un tel protocole favorisera la reproductibilité intra et inter-laboratoire de la procédure de préparation des échantillons en immunopeptidomique et devrait trouver des applications dans la recherche fondamentale et clinique.

# **CHAPITRE 3 – Logiciels pour évaluation rapide de la spécificité et qualité générale des immunopeptidomes identifiés**

## **3.1 - Problématiques de contrôle de la qualité (CQ) en immunopeptidomique**

Très peu d'ouvrages décrivent le développement d'approches CQ en immunopeptidomique[58, 142]. Le CQ deviendra d'une importance critique dans l'immunopeptidomique basée sur la spectrométrie de masse afin de se conformer aux réglementations pharmaceutiques pour les applications cliniques réelles[58]. Pour diverses raisons, le développement d'outils logiciels de CQ n'a pas été une priorité en immunopeptidomique au cours des dernières années. La raison principale étant que la plupart des études immunopeptidomiques doivent encore traiter des ensembles de données à petite échelle et que la qualité de ces ensembles de données peut être évaluée manuellement à l'aide des algorithmes de regroupement de séquences peptidiques et de liaison de peptides du CMH actuellement disponibles. Cependant, avec le développement récent de technologies automatisées pour des protocoles d'immunoprécipitation du CMH robustes et à haut débit, il deviendra plus commun d'avoir à analyser des ensembles de données immunopeptidomiques à l'échelle de plusieurs giga-octets qui ne se prêteront pas à une analyse manuelle des données [143].

À cet égard, MHCvizPipe (MVP) est le premier outil-logiciel qui se concentre spécifiquement sur le contrôle de la qualité des échantillons en immunopeptidomique basée sur la SM. La première version de MVP peut être facilement appliquée pour évaluer la qualité des données provenant d'ensembles d'échantillons de petite à moyenne taille.

### **3.1.2 - MHCvizPipe : L'avenir de l'analyse d'immunopeptidomique**

La qualité des données est la pierre angulaire de recherches exigeant répétabilité et reproductibilité. À cet égard, il est important d'évaluer et de contrôler la qualité des données immunopeptidomiques obtenues par SM. Dans les domaines de la génomique

et de la protéomique, l'importance du contrôle et de l'assurance de la qualité est reconnue depuis longtemps et différents types d'échantillons de contrôle de la qualité ainsi que des solutions informatiques de CQ ont été largement développés et appliqués au fil des ans[144]. À l'inverse pour le cas de l'immunopectidomique, les échantillons et les outils logiciels de contrôle de la qualité spécialisés pour les peptides associés aux CMH de classe I et II restent pratiquement non-documentés dans la littérature[144-146]. Les logiciels actuels permettant d'évaluer la qualité et la spécificité des ensembles de données en immunopectidomiques générés par SM comprennent des algorithmes de prédiction de la liaison des peptides du CMH et des outils de « clustering » tels que NetMHCpan[47, 51], MHCFlurry[147, 148], GibbCluster[46] et MoDEC[149]. Bien qu'ils soient largement utilisés en protéomique, ces algorithmes n'ont pas été conçus spécialement pour le contrôle de la qualité de l'immunopectidomique par SM et ils ne peuvent traiter qu'un seul échantillon à la fois et nécessitent généralement une évaluation supplémentaire ou d'autres manipulations des données par l'utilisateur. De telles procédures sont relativement longues et sujettes aux erreurs, ce qui n'est pas viable dans le cas d'études à grande échelle, comme récemment signalé[87, 150]. Par conséquent, le développement de logiciels automatisés pour une évaluation rapide et simultanée de la qualité de multiples ensembles de données générés par la SM reste à développer. Pour ces raisons, le développement de systèmes de contrôle d'assurance et de contrôle de la qualité capables de détecter les problèmes liés aux techniques d'isolation ou à la nature des échantillons sont importants pour une gestion adéquate et une harmonisation des données.

Dans cette optique, Kovalchik, K.A et al (2021) ont créé MhcVizPipe [117], le premier outil logiciel de contrôle de qualité semi-automatisé qui permet une évaluation rapide et simultanée de multiples ensembles de données immunopectidomiques du CMH de classe I et II générés par SM. MVP fournit une évaluation visuelle graphique rapide de la qualité, de la composition et de la spécificité du CMH de l'échantillon, ce qui permet d'accélérer considérablement le processus décisionnel de contrôle de qualité en vue de l'interprétation des données. MVP tire profit de l'utilisation d'algorithmes

immunopeptidomiques bien établis (NetMHCpan, NetMHCIIpan et GibbsCluster) et génère à une vitesse sans précédent des rapports organisés et faciles à comprendre sous format HTML. Les paramètres de contrôle de la qualité rapportés sont la longueur des peptides, le nombre de peptides pouvant se fixer au CMH, la distribution des peptides par allèle du CMH et les principaux motifs de fixation des niches peptidiques du CMH. Ces rapports peuvent être exportés et consultés sur la plupart des ordinateurs équipés d'un navigateur Web moderne. Avec sa facilité d'utilisation, MVP pourrait être utilisé dans des laboratoires spécialisés en immunopeptidomique autant que dans des laboratoires novices qui élargissent leurs champs d'intérêts vers l'immunopeptidomique et ce, dans le but d'uniformiser les méthodes d'analyses et les données en immunopeptidomique générés par SM.

L'objectif de ce chapitre (objectif 3) est d'utiliser le logiciel MHCvizPipe comme outil de contrôle de la qualité pour l'évaluation rapide de la spécificité et reproductibilité d'immunopeptidomes isolés avec les techniques observées au chapitre 2. Étant donné que ces échantillons sont analysés à travers la même plateforme expérimentale, on s'attend à voir une haute reproductibilité des résultats.

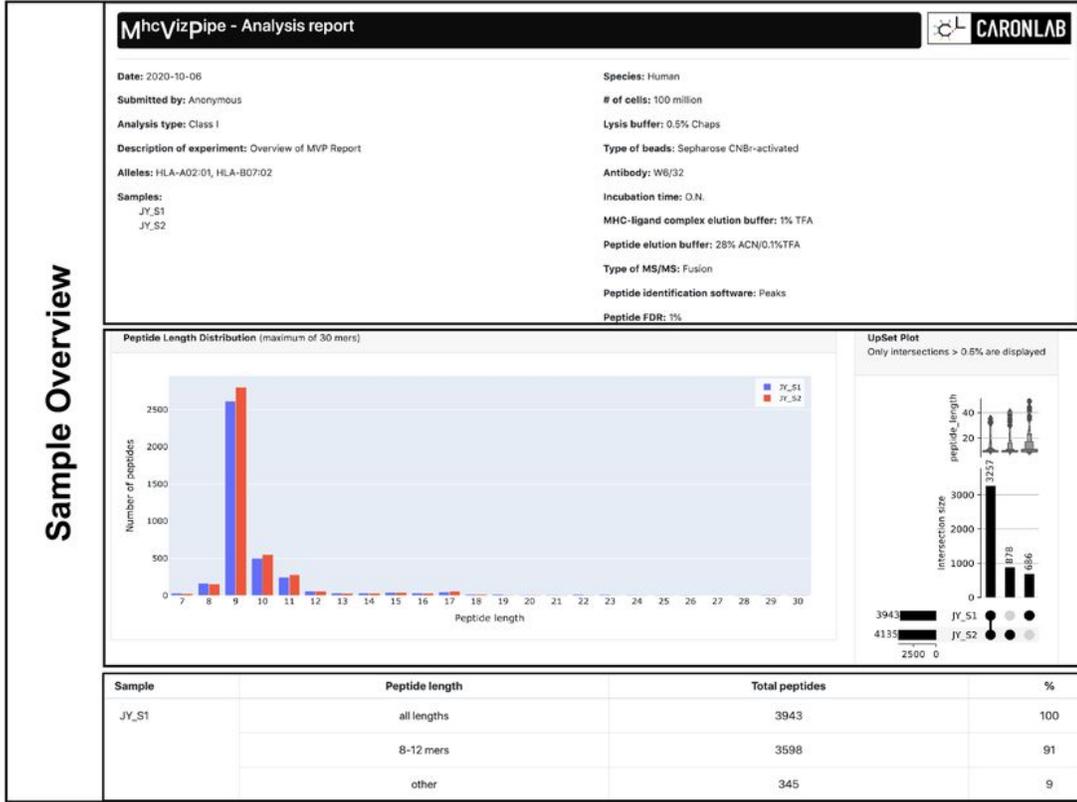
### **3.2 – Méthode : Composantes utilisées dans MHCvizpipe**

MVP tire profit des outils bioinformatiques NetMHCpan, NetMHCIIpan et GibbsCluster et fabrique des rapports HTML portables pour l'interprétation des données d'immunopeptidomique par spectrométrie de masse grâce à des bibliothèques de visualisation Python. En plus de NetMHCpan, NetMHCIIpan et GibbsCluster, il utilise les bibliothèques Python suivantes : Plotly, PlotlyDash, Dash Bootstrap Components, Pandas, Numpy, Dominate, UpSetPlot, Seaborn, Waitress et Plotly Logo. L'interface de MVP est construite à l'aide de la bibliothèque PlotlyDash et fonctionne comme une application web locale.

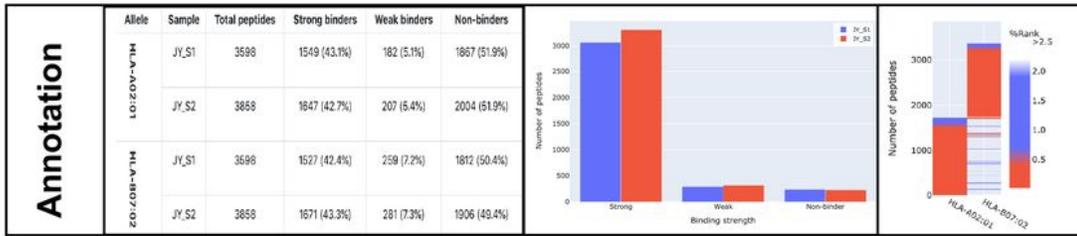
### 3.2.1 - NetMHCpan and NetMHCIIpan

Avant l'analyse avec NetMHCpan ou NetMHCIIpan, des listes de peptides sont sélectionnés avec les critères suivants: pour les données concernant des peptides du CMH de classe I; les longueurs sont limitées entre 8 et 12 acides aminés inclusivement ; pour la classe II, les longueurs sont limitées entre 9 et 22 acides aminés inclusivement. Ces listes sont analysées à l'aide de NetMHCpan ou NetMHCIIpan pour obtenir des prédictions de liaison (rang en pourcentage du ligand élué) utilisées pour annoter les peptides selon leur score de prédiction de l'affinité décrit comme suit : pour le CMH - I, le rang des liants forts est plus petit ou égal à 0,5, celui des liants faibles est entre 0,5 et 2,0 et celui des non liants est  $> 2,0$  ; pour le CMH - II, le rang des liants forts est plus petit ou égal à 2,0, celui des liants faibles est entre 2 et 10 et celui des non liants est  $> 10$ . Les résultats sont présentés sous forme de tableau, de diagramme à barres et de carte thermique (**Fig. 13**). Le diagramme à barres (**Fig. 13 B**) montre le nombre de peptides en fonction de la force de liaison (par exemple, un peptide qui a une faible affinité pour un allèle et une forte affinité pour un autre est considéré comme un liant fort). Une « *heat-map* » est créée pour représenter le score d'affinité des différents peptides pour chaque allèle du CMH retrouvé dans le modèle cellulaire (**Fig. 13 B, extrémité droite**). Afin d'éviter que l'ordre ne soit faussé par des peptides à faible score, toutes les scores de prédiction d'affinité supérieurs à 2,5 pour la classe I ou à 12 pour la classe II sont fixés à 2,5 ou 12, respectivement. Les ligands sont définis de telle sorte que le rouge représente approximativement les ligands forts, le bleu les ligands faibles et le jaune les non liants.

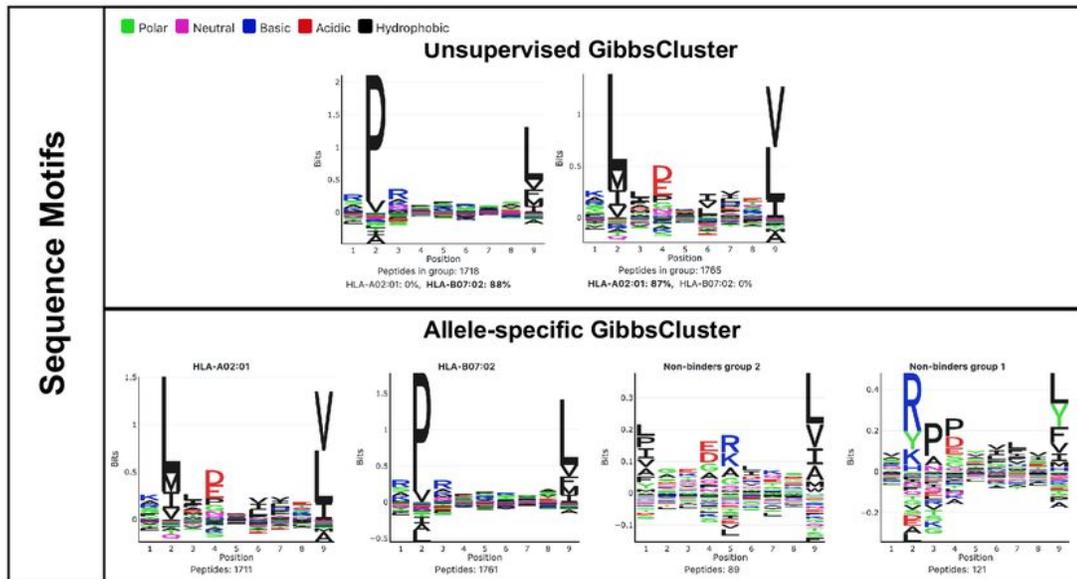
A



B



C



**Figure 13 :** Aperçu du rapport MHCvizPipe. Deux réplicats d'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules JY ont été utilisées pour présenter ce rapport HTML : (A). Aperçu visuel de la section "Sample Overview" du rapport MVP. Un tableau contenant le nombre de peptides est présenté sous le graphique de distribution de la longueur des peptides (à gauche) et le graphique UpSet (à droite). (B). Section résultats d'annotation obtenu à partir de NetMHCpan (I ou II). À gauche : tableau descriptif de la proportion du nombre total, des peptides du CMH forts, faibles et non liants. Au milieu : Graphique montrant le nombre de peptides en fonction de l'affinité de liaison. A droite : carte thermique montrant le nombre de peptides en fonction de l'affinité de liaison pour chaque allèle analysé. (C). Représentation visuelle de la section Motifs de séquence de l'analyse GibbsCluster. Adapté de Kovalchik, K et al (2021) [117]

- En tant que référence, l'article "MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- to Large- Scale Immunopeptidome Data Sets" a été ajouté en annexe 4 [117]

### 3.2.2 – GibbsCluster

MVP exécute deux rondes d'analyses via GibbsCluster appelées "Unsupervised GibbsCluster" et "Allele-Specific GibbsCluster". L'analyse non supervisée est un GibbsCluster standard exécuté en utilisant tous les peptides du sous-ensemble décrit plus tôt. L'analyse « *Allele-specific GibbsCluster* » dépend des résultats de NetMHCpan ou NetMHCIIpan. Un sous-ensemble de peptides est créé pour chaque allèle. Tous les peptides d'un sous-ensemble sont des liants forts ou faibles pour l'allèle respectif. Un sous-ensemble supplémentaire est créé dans lequel tous les peptides non liants pour aucuns allèles sont combinés. Tout sous-ensemble contenant moins de 20 peptides est écarté. Les résultats sont présentés dans l'onglet « *Allele-Specific GibbsCluster* », où un motif est affiché pour chaque allèle présent dans l'échantillon, et jusqu'à 5 motifs pour les peptides non liants. L'analyse ayant le score KLD (*Kullback-Leibler Distance*) le plus élevé est celle qui sera présenté dans le rapport. Tous les groupes de peptides présentés dans la section "*Sequence Motifs*" sont visualisés à l'aide de la bibliothèque PlotlyLogo.

Pour réduire le temps d'analyse global, MVP parallélise l'utilisation de NetMHCpan, NetMHCIIpan et GibbsCluster. GibbsCluster, offre le multitraitement des données, mais comme NetMHCpan et NetMHCIIpan ne sont pas en mesure de traiter plusieurs ensembles de données de manière simultanée, MVP divise les groupes de peptides en plus petites listes qui sont analysées simultanément dans plusieurs onglets qui exécutent les logiciels, et ce, à plusieurs reprises, de manière simultanée. Les résultats sont ensuite combinés en une seule liste.

### **3.2.3 - MVP annote les peptides à l'aide de NetMHCpan/IIpan**

NetMHCpan (v.4.0 ou 4.1) et NetMHCIIpan (v.4.0) sont des composants essentiels du logiciel MVP pour évaluer les peptides de classe I et II respectivement. Les scores qui en résultent définissent des ligands forts ( $\%Rank < 0,5$ ), des ligands faibles ( $2,0 < \%Rank > 0,5$ ) ou des non liants ( $\%Rank > 2,0$ ). Ces scores sont ensuite traités pour annoter les peptides de classe I et II à leur allèle CMH respectif, comme décrit par Caron, E et Shao, W[97, 110]. En d'autres termes, ces scores peuvent servir d'indice de spécificité de l'échantillon pour les allèles attendus selon le modèle d'étude utilisé.

### **3.2.4 - MVP génère un rapport HTML détaillé**

Le rapport HTML généré par l'interface utilisateur graphique du MVP contient trois sections principales : Aperçu de l'échantillon, résultats d'annotation avec cartes thermiques et séquence des motifs de liaisons.

### **3.2.5 – Aperçu de l'échantillon**

La première section du rapport HTML contient un graphique de distribution des longueurs de peptide (jusqu'à 30 acides aminés), un tableau descriptif du nombre de peptides comprenant le nombre et la proportion de peptides correspondant à la longueur attendue en fonction de la classe de CMH sélectionnée, et un graphique « UpSet » [151] montrant le nombre de peptides uniques et partagés entre plusieurs échantillons.

### 3.2.6 - Résultats d'annotation avec cartes thermiques

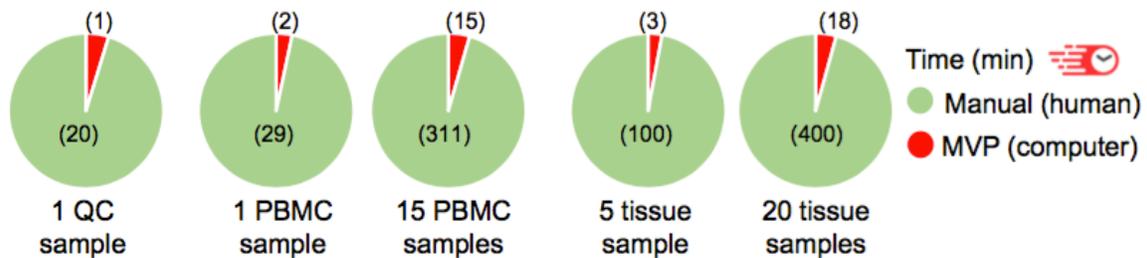
MVP utilise les scores issus de la méthode de notation « *Eluted ligand* » des logiciels NetMHC pour générer des histogrammes et cartes thermiques illustrant la proportion de peptides purifiés par immuno-affinité qui sont en mesure de se lier aux molécules du CMH, ce qui permet de confirmer la spécificité des peptides pour les allèles du CMH des cellules desquelles ils ont été isolés [51]. Ces graphiques donnent une image claire de la qualité et de la composition des échantillons et sont judicieux pour évaluer la spécificité de l'échantillon, tant pour les échantillons de contrôle de la qualité que pour les échantillons biologiques analysés. Les peptides de classe I élués des cellules JY et EL4 utilisés en tant que contrôles positifs représentent des normes de contrôle de qualité biologique pour évaluer les performances des instruments et de la préparation des échantillons. Si un échantillon ne répond pas aux critères d'évaluation de la qualité prédéfinis pour la longueur ou l'association avec un ou plusieurs allèles spécifiques du CMH, ce rapport avertit l'utilisateur qu'une enquête plus approfondie est nécessaire avant de poursuivre l'interprétation des données.

### 3.2.7 - Séquence des motifs de liaisons

La troisième section du rapport HTML se concentre sur l'identification et la visualisation des motifs de séquence qui sont enrichis dans les échantillons analysés. Les motifs de séquence issus de l'analyse GibbsCluster sont présentés de deux manières. Premièrement, l'onglet non supervisé affiche les motifs les plus importants ainsi que le pourcentage de peptides associés à chaque allèle. Ensuite, l'onglet « *allele-specific* » affiche les motifs peptidiques générés à partir des liants forts et faibles du CMH pour chaque allèle, ainsi que les motifs des peptides prédits comme étant des liants du CMH non attribués. MVP génère en quelques clics un rapport HTML complet pour évaluer la qualité des données immunopeptidomiques générées par la spectrométrie de masse.

L'un des principaux avantages du traitement des données immunopeptidomiques par MVP est la rapidité avec laquelle les rapports HTML sont générés pour un ou plusieurs échantillons, accélérant ainsi le processus d'analyse et de génération de figures/tableaux

par ~22 fois en moyenne (**Fig. 14**). MVP représente le premier logiciel de contrôle de la qualité en immunopeptidomique basée sur la spectrométrie de masse et offre une vitesse sans précédent pour la visualisation et l'évaluation rapide de la qualité des données immunopeptidomiques.



**Figure 14** : Performance à haute vitesse du logiciel MVP pour le CQ en immunopeptidomique basée sur la SM. Diagrammes circulaires montrant les estimations de temps pour traiter et analyser les ensembles de données sélectionnés manuellement (en vert) ou en utilisant MVP (en rouge). Les ensembles de données testés sont : (de gauche à droite) 1 échantillon QC (JY), 1 échantillon PBMC, 15 échantillons PBMC, 5 échantillons de tissus de souris et 20 échantillons de tissus de souris. Adapté de Kovalchik, K et al. [117] (Référer à l'annexe 4 pour plus d'informations)

### 3.3 - Résultats

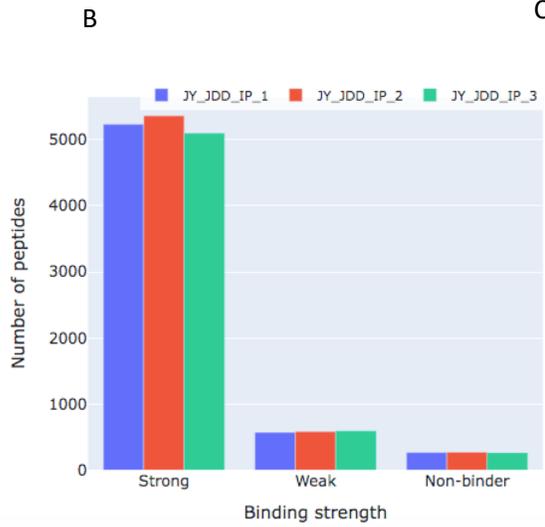
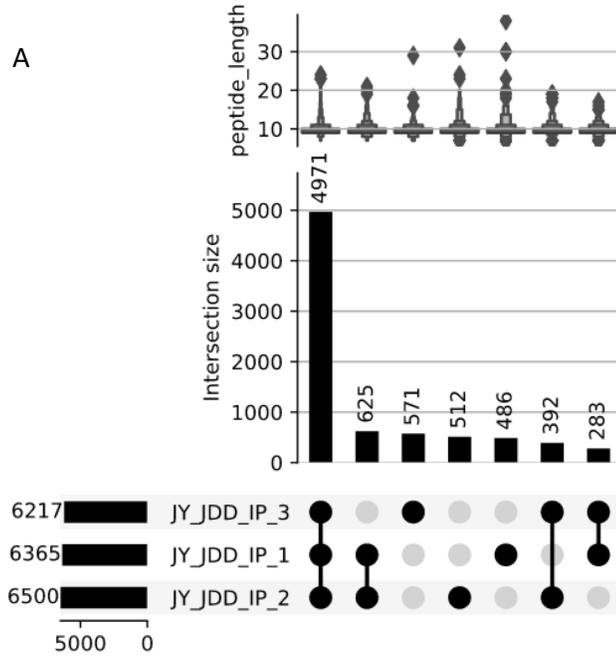
#### 3.3.1 - Enjeux de recherches reproductibles en immunopeptidomique par spectrométrie de masse

Pour permettre le déploiement de l'immunopeptidomique, il est essentiel d'établir et de confirmer la reproductibilité des résultats d'isolation de peptides et des analyses de spectrométrie de masse et ce, entre les différents usagers, instruments, institutions et logiciels d'analyses.

#### 3.3.2 - Reproductibilité intra-utilisateur

Dans le but de déterminer la reproductibilité intra-utilisateur dans les procédures d'isolation de peptides, nous avons analysé, à l'aide de MVP, la nature des peptides issus de trois répétitions de procédure d'immuno-précipitation par affinité sur  $100 \times 10^6$  cellules

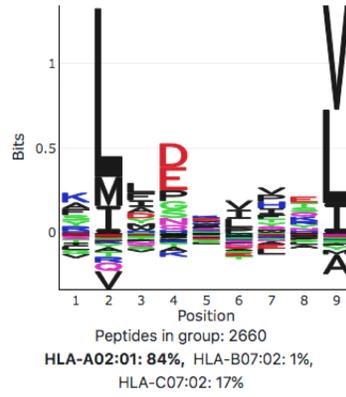
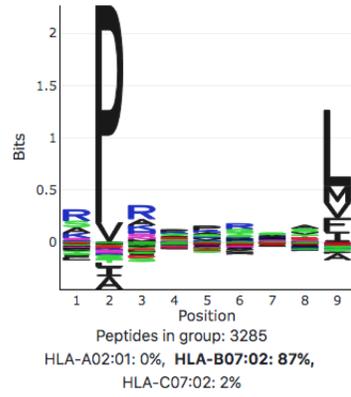
JY à l'aide de l'anticorps W6/32 (**Fig. 15**). Les données brutes ont été analysés dans un même rapport HTML et nous permettent de facilement visualiser les variations entre les trois répétitions. 6365, 6500 et 6217 peptides ont été identifiés respectivement dans les expérimentations 1,2 et 3, représentant un maximum de coefficient de variation de 4.4% entre les expériences 2 et 3(**Fig.15.A**). 4971 peptides ont été identifiés dans les trois répétitions, représentant 76,5% du nombre maximal de peptides retrouvés dans une répétition. De plus, parmi les peptides qui ne sont pas partagés par les 3 répétitions, 625 (9,6%) des peptides se retrouvent dans les expériences 1 et 2, 392 (6%) dans les expériences 2 et 3 et 283 peptides (4,3%) dans les expériences 1 et 3. Il est donc possible de conclure que au maximum, 86,1% des peptides totaux se retrouvent dans les répétitions 1 et 2(**Fig.15.A**), ce qui démontre une reproductibilité surprenante compte tenu que l'acquisition a été effectué en mode DDA ou une grande proportion des ions ne sont pas sélectionnés pour être fragmentés[44]. Une moyenne de 4.33% des peptides sont considérés comme étant des contaminants car ils ne possèdent pas le nombre de peptides adéquats pour être portés sur les molécules du CMH de classe 1(**Fig.15.B**). Une moyenne de 37,5% des peptides à travers les répétitions sont prédit pour être fortement liés aux molécules de l'allèle A02:01 avec un taux de variation maximum de 1% entre les expériences 1 et 2. En moyenne, 46,7% des peptides sont prédit pour avoir une forte interaction avec les molécules de l'allèle B07:02 avec une variation maximale de 1,3% entre les expériences 1 et 3(**Fig.15.C**). Toutes les expériences ont des résultats cohérent au niveau des motifs de liaison. Les peptides ayant une forte affinité pour les allèles A02:01 et B07:02 portent majoritairement les acides aminés L, V et P, L respectivement aux positions d'ancrage 2 et 9(**Fig.15.D**).



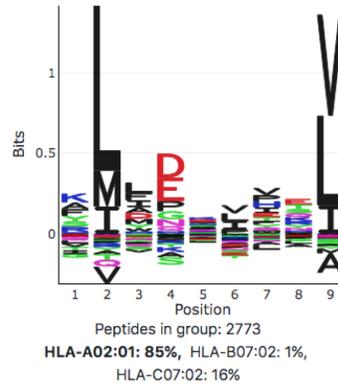
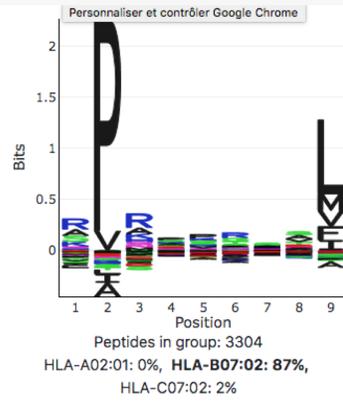
**C**

Allele	Sample	Total peptides	Strong binders	Weak binders	Non-binders
HLA-A*02:01	JY_JDD_IP_1	6074	2250 (37.0%)	323 (5.3%)	3501 (57.6%)
	JY_JDD_IP_2	6219	2363 (38.0%)	338 (5.4%)	3518 (56.6%)
	JY_JDD_IP_3	5955	2248 (37.7%)	342 (5.7%)	3365 (56.5%)
HLA-B*07:02	JY_JDD_IP_1	6074	2878 (47.4%)	439 (7.2%)	2757 (45.4%)
	JY_JDD_IP_2	6219	2908 (46.8%)	450 (7.2%)	2861 (46.0%)
	JY_JDD_IP_3	5955	2743 (46.1%)	451 (7.6%)	2761 (46.4%)
HLA-C*07:02	JY_JDD_IP_1	6074	566 (9.3%)	1427 (23.5%)	4081 (67.2%)
	JY_JDD_IP_2	6219	573 (9.2%)	1466 (23.6%)	4180 (67.2%)
	JY_JDD_IP_3	5955	551 (9.3%)	1363 (22.9%)	4041 (67.9%)

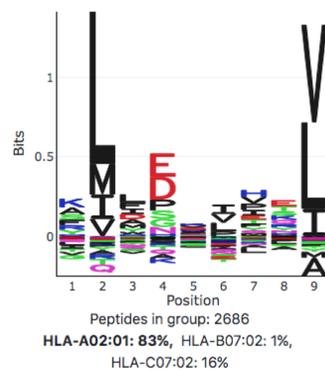
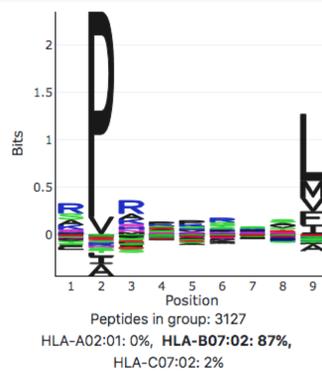
D JY\_JDD\_IP\_1 (peptides used: 6074, outliers: 129)



JY\_JDD\_IP\_2 (peptides used: 6219, outliers: 142)



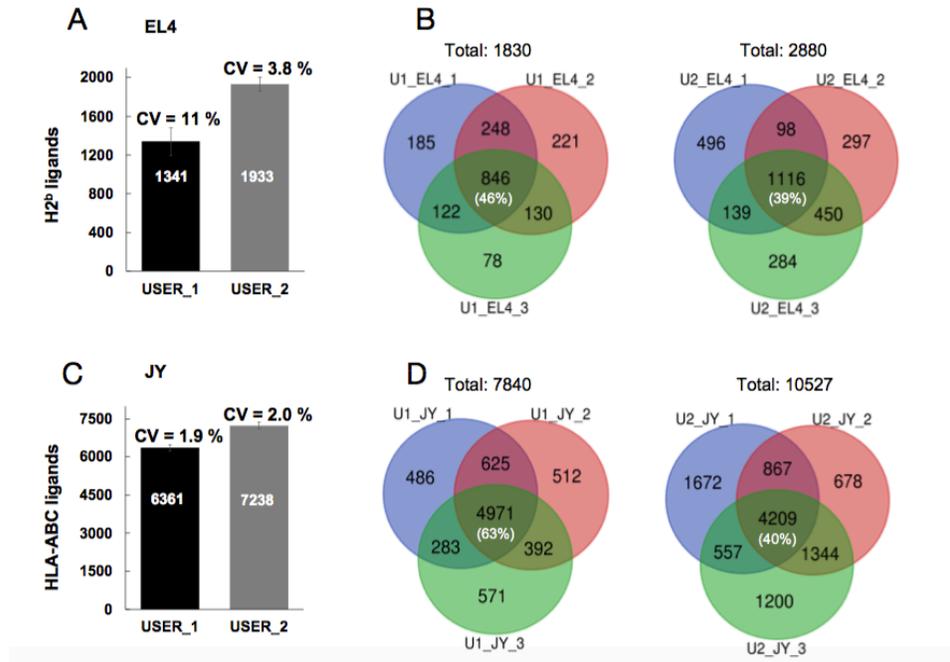
JY\_JDD\_IP\_3 (peptides used: 5955, outliers: 142)



**Figure 15** –Analyse de trois réplicats d'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules JY par le même utilisateur. (A) Section « *Plotly Logo* » affichant les peptides communs retrouvés dans les 3 expériences. (B) Section « *Binding Affinities* » démontrant la proportion des peptides ayant une affinité forte (*strong*), faibles (*weak*) ou sans affinités (*non-binder*) à travers les 3 réplicats biologiques. (C) Section « *Annotation Results* », où la prédiction d'affinité des peptides est analysé pour chaque allèles du CMH-1 retrouvées dans cette lignée cellulaire. (D) Représentation visuelle de la section « *Sequence Motif* » issue de l'analyse GibbsCluster.

### 3.3.3 - Reproductibilité inter-utilisateur/intra-établissement

Afin d'évaluer la reproductibilité des résultats du protocole d'isolation des peptides du CMH, les résultats de deux membres du laboratoire ayant chacun effectués trois réplicats biologiques ont été comparés. Pour l'utilisateur 1, le nombre moyen de peptides spécifiques de MHC- I moyen détectés à partir des cellules EL4 et JY était de 1341 et 6361, respectivement ; pour l'utilisateur 2, 1933 et 7238. Le coefficient de variation moyen du nombre de peptides détectés dans trois réplicas biologique différents varie de 1,9 à 11 % (**Fig. 16 A, C**). Bien que les coefficients de variation pour le nombre de peptides identifiés dans les trois expériences différentes soient relativement faibles, l'identité des peptides détectés varie considérablement (**Fig. 16 B, D**). Les diagrammes de Venn en figure 17 montrent que la proportion de peptides qui ont été détectés dans les trois réplicas biologiques est modérée à élevée, allant de 39 % (utilisateur 2, cellules EL4) à 63 % (utilisateur 1, cellules JY (**Fig. 16 B, D**)[135]. Ceci démontre pleinement les limitations des méthodes d'acquisition de données de type DDA ou la nature courte des peptides associés aux CMH ainsi que la faible proportion des peptides contenant des acides aminés basiques empêchent une ionisation efficace et génèrent des spectres MS2 peu prévisibles.



**Figure 16** : Identification du nombre de peptides du CMH de classe I isolés à partir de cellules JY (A) et EL4 (C). (B, D) Diagramme Venn illustrant les peptides retrouvés dans les 3 réplicats biologique des deux utilisateurs. Adapté de Sirois, I, Despault, J et al [135]

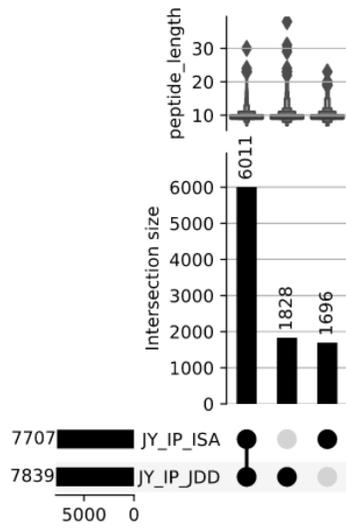
Également afin de déterminer la reproductibilité inter-utilisateurs au sein d'un même laboratoire, nous avons utilisé MVP pour générer un rapport comparant les résultats de deux utilisateurs qui ont utilisés le protocole d'isolation de peptides du CMH-I avec billes activées sépharose CNBr sur  $100 \times 10^6$  cellules JY avec l'anticorps W6/32 (**Fig.17**) ainsi que sur  $100 \times 10^6$  cellules EL4 avec l'anticorps M1 (**Fig. 18**). Les résultats démontrent une nette reproductibilité à tous les niveaux. Au sein des cellules JY, l'utilisateur 1 (JDD) à isolé 7839 peptides comparativement à 7707 pour l'utilisateur 2 (ISA), représentant un très faible taux de variation de 1.7%. De ces peptides, 6011 (76%) ont été identifiés par les deux utilisateurs (**Fig.17.A, C**). 374 et 386 peptides n'étant pas de longueur requise pour être portés par une molécule du CMH de classe 1 ont été identifiés respectivement par les utilisateurs 1 et 2(**Fig.17.C**). Ceci représente uniquement 4,8 et 5% respectivement pour les utilisateurs 1 et 2 et démontre l'efficacité et la spécificité de la procédure d'isolation des peptides. Des résultats similaires sont observés au niveau de la distribution des peptides selon les divers allèles du CMH trouvés au sein

des cellules JY. 37,1 et 35,3% des peptides sont des ligands fort pour les molécules de l'allèle A02:01, exactement 46% des peptides isolés par les deux utilisateurs sont prédit pour être des ligands forts pour l'allèle B07:02 et 9.1 et 8,7% pour l'allèle C07:02 (**Fig.17.B**). Les motifs de liaisons affichent aussi une grande homologie. Dans les deux cas, les peptides prédit pour avoir une grande affinité pour les allèles A02:01 et B07:02 portent majoritairement les acides aminés L, V et P, L aux positions d'ancrage 2 et 9 (**Fig. 17.D**).

Au niveau des cellules EL4 (**Fig. 18**), des résultats tout aussi cohérent sont observés entre les utilisateurs 1 et 2. 1752 peptides ont été isolés par l'utilisateur 1(JDD) et 1767 peptides par l'utilisateur 2(ISA), représentant un taux de variation de 0,8%. De ces peptides, 1308 sont identifiés par les deux utilisateurs (**Fig. 18.A**) (74%), ce qui représente pratiquement la même proportion que dans l'expérimentation avec les cellules JY (**Fig. 19.A**). Au niveau de la distribution des peptides selon les divers allèles du CMH trouvés au sein des cellules EL4, 4,1% de variation est observé au niveau des peptides prédit pour s'attacher de manière forte à l'allèle H-2 Kb tandis que 3,7% de différence est observée pour l'allèle H-2 Db(**Fig. 19.B**). La composition des acides aminés aux positions d'ancrage est aussi homogène dans les deux expérimentations, soit F et L pour H-2 Kb et N et L pour H-2 Db(**Fig. 19.D**).

Malgré que ces expériences aient été réalisées dans un même laboratoire, le nombre de peptides communs identifiés, les proportions de ligands forts ainsi que les motifs de liaisons, autant dans les expériences intra ou inter utilisateurs, démontrent une reproductibilité accrue dans la nature et composition des peptides isolés par immunopurification par affinité.

A



B

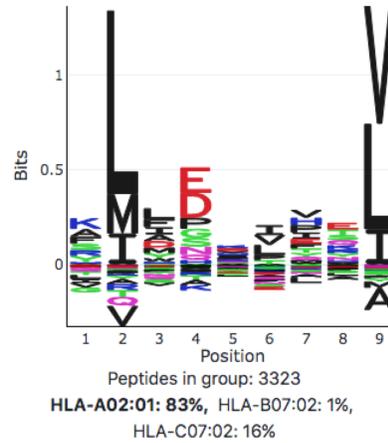
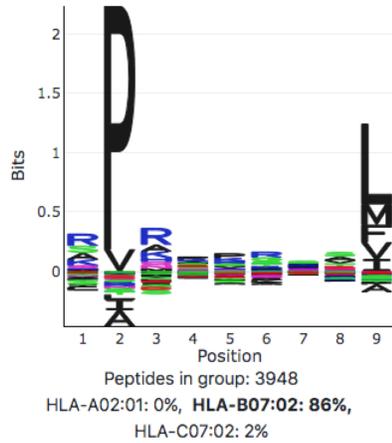
Allele	Sample	Total peptides	Strong binders	Weak binders	Non-binders
HLA-A*02:01	JY_IP_JDD	7465	2773 (37.1%)	420 (5.6%)	4272 (57.2%)
	JY_IP_ISA	7321	2587 (35.3%)	428 (5.8%)	4306 (58.8%)
HLA-B*07:02	JY_IP_JDD	7465	3431 (46.0%)	550 (7.4%)	3484 (46.7%)
	JY_IP_ISA	7321	3364 (46.0%)	597 (8.2%)	3360 (45.9%)
HLA-C*07:02	JY_IP_JDD	7465	682 (9.1%)	1708 (22.9%)	5075 (68.0%)
	JY_IP_ISA	7321	639 (8.7%)	1640 (22.4%)	5042 (68.9%)

C

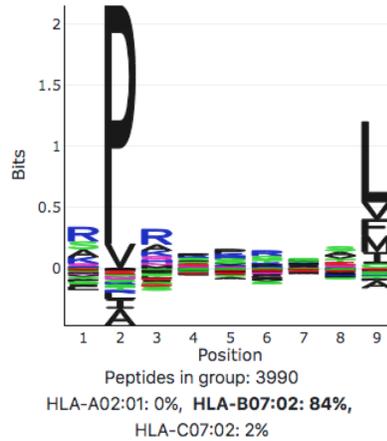
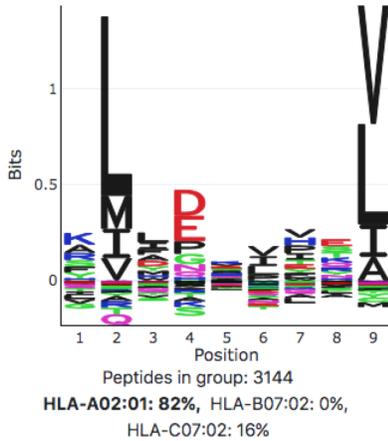
Sample	Peptide length	Total peptides	%
JY_IP_JDD	all lengths	7839	100
	8-12 mers	7465	95
	other	374	5
JY_IP_ISA	all lengths	7707	100
	8-12 mers	7321	95
	other	386	5

D

JY\_IP\_JDD (peptides used: 7465, outliers: 194)

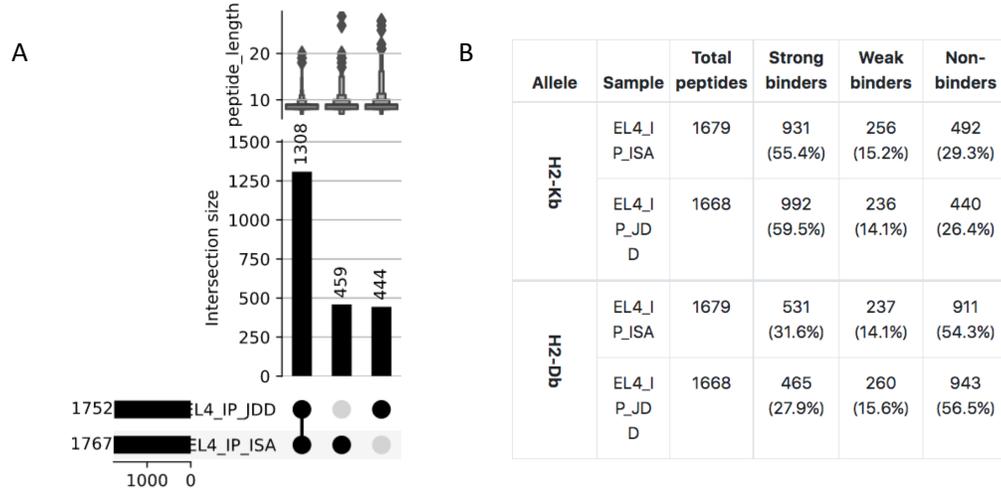


JY\_IP\_ISA (peptides used: 7321, outliers: 187)



**Figure 17** : Rapport MVP suite à l'analyse de l'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules JY par deux utilisateurs (JDD, ISA) d'un même laboratoire. (A) Section « *Plotly Logo* » affichant les peptides communs retrouvés dans les 3 expériences. (B) Section « *Annotation Results* », où la prédiction d'affinité des peptides est analysé pour chaque allèles du CMH-1 retrouvées dans cette lignée cellulaire. (C) Section « *Overview1* » où l'on peut observer la quantité des peptides qui à la taille requise pour interagir avec les molécules de

CMH-1. (D) Représentation visuelle de la section « *Sequence Motif* » issue de l'analyse GibbsCluster.

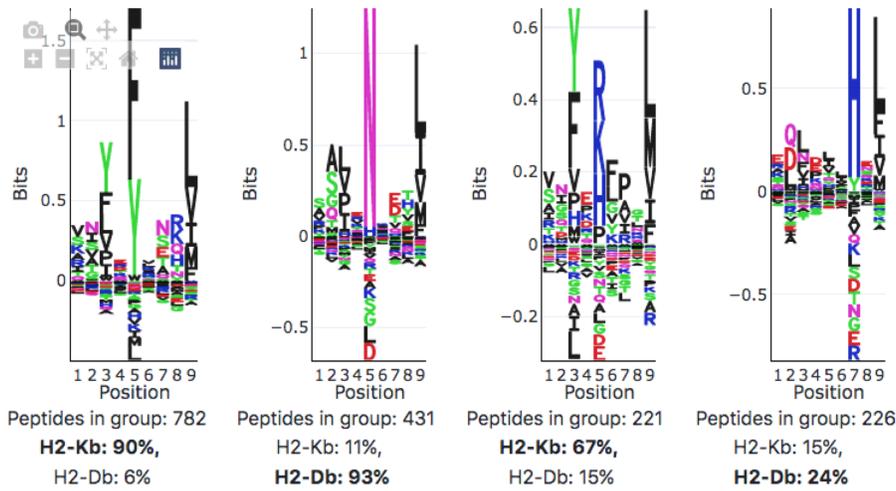


**C**

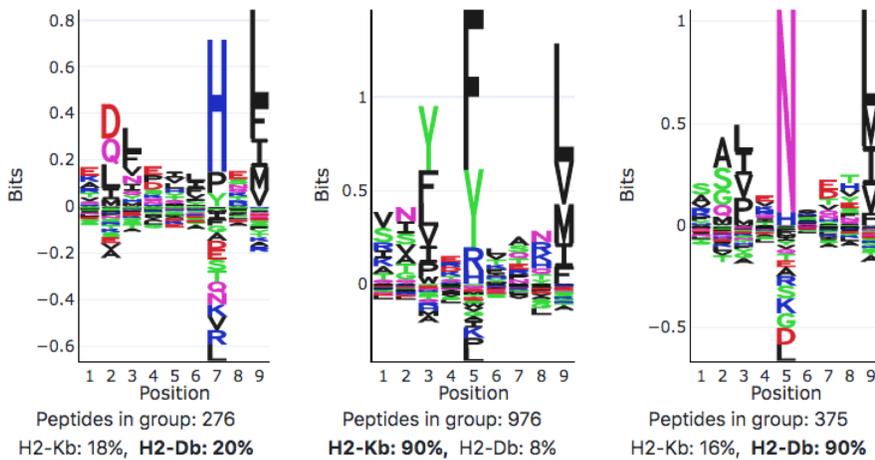
Sample	Peptide length	Total peptides	%
EL4_IP_ISA	all lengths	1767	100
	8-12 mers	1679	95
	other	88	5
EL4_IP_JDD	all lengths	1752	100
	8-12 mers	1668	95
	other	84	5

D

EL4\_IP\_ISA (peptides used: 1679, outliers: 19)



EL4\_IP\_JDD (peptides used: 1668, outliers: 41)



**Figure 18** : Rapport MVP suite à l'analyse de l'immunopeptidomes du CMH de classe I isolés à partir de 100 millions de cellules EL4 par deux utilisateurs d'un même laboratoire (JDD, ISA). (A) Section « *Plotly Logo* » affichant les peptides communs retrouvés dans les 3 expériences. (B) Section « *Annotation Results* », ou la prédiction d'affinité des peptides est analysé pour chaque allèles du CMH-1 retrouvées dans cette lignée cellulaire. (C) Section « *Overview* » où l'on peut observer la quantité des peptides qui à la taille requise pour interagir avec les molécules de

CMH-1. (D) Représentation visuelle de la section « *Sequence Motif* » issue de l'analyse GibbsCluster.

### **3.4 – Discussion : L’avenir des analyses des ligands du MHC par spectrométrie de masse**

Les peptides anormaux présentés par les molécules du complexe majeur d'histocompatibilité sont très prometteurs pour le traitement du cancer, des maladies infectieuses, de l'auto-immunité et des allergies en raison de leur rôle central dans la surveillance immunitaire et l’activation des cellules T[13, 30, 60, 143].

Malgré des développements considérables dans les prédictions *in-silico* des complexes peptides-CMH au cours des dernières années, les observations *in-vitro* révèlent continuellement des écarts entre les peptides prédits de manière computationnelles et les peptides générés de manière biologique par la machinerie cellulaire. Le séquençage des peptides par SM permet de contourner les limites comprises dans les logiciels de prédictions. L'identification par LC-MS/MS permet de révéler des modifications post-traductionnelles, l'épissage alternatif et la présentation par des allèles de CMH mal caractérisés tel que ceux mentionnés plus tôt (HLA-C). Certains peptides ayant une faible interaction au CMH associés à des états pathologiques pourraient échapper à une prédiction de l’affinité effectuée de manière computationnelle[143]. Comme mentionné à de nombreuses reprises lors de cet ouvrage, les méthodes d’isolation des peptides du CMH sont restés relativement inchangés et manquent d’efficacité, non-seulement au niveau de la reproductibilité, mais présentent aussi des limitations concernant le nombre d’échantillons pouvant être analysés de manière simultanée[57, 60, 98, 101].

Pour que l'immunoprécipitation par affinité des peptides du CMH devienne un outil fiable et normalisé entre les institutions, il sera primordial d’établir un protocole standardisé capable de traiter de nombreux échantillons en même temps. Lichao Zhang et al (2021) ont décrit un nouveau protocole pour la purification des peptides du CMH utilisant la plateforme AssayMAP Bravo d'Agilent Technologies qui automatise presque entièrement l’analyse de lysat cellulaire en utilisant un dispositif automatisé de

manipulation de liquides sur des plaques à 96 puits. Un tel protocole d'immuno-précipitation du CMH à haut débit permettrait des améliorations extraordinaires en termes de débit, de vitesse, de sensibilité et reproductibilité, et ce, intra/inter utilisateurs à travers multiples institutions, dans le but ultime d'harmoniser les protocoles d'extractions des peptides du CMH [143].

## **CHAPITRE 4 - Reproductibilité inter-établissement, application des techniques d'isolation et utilisation des logiciels pour l'analyse de l'immunopeptidome de cellules permissives à l'infection au Coronavirus.**

### **4.1 - La reproductibilité inter-établissement dans l'étude de l'immunopeptidomique**

L'atteinte de l'objectif de reproductibilité des données au sein des membres d'une institution utilisant le même instrument d'analyse est un premier pas vers une harmonisation des résultats en immunopeptidomique. Cependant, pour obtenir des résultats homogènes en analysant des ensembles de données de taille considérable avec des résultats reproductibles, il faut intégrer efficacement les données recueillies par de multiples instruments provenant de différentes institutions, et ce, sur de longues périodes de temps. Il est donc impératif de connaître l'impact et la nature des facteurs qui affectent la reproductibilité de l'identification des peptides par SM, et de développer des techniques d'analyse des données pour optimiser l'intégration de ces données de manière robuste et consistante à travers le temps entre divers instruments.

L'objectif de cette section (objectif 4) est d'appliquer les méthodes et outils de contrôle de qualité des objectifs 1, 2 et 3 pour évaluer la reproductibilité de l'analyse de l'immunopeptidome de cellules A549-ACE2/TMPRSS2 inter-établissement. Pour ce faire, nous avons tenté de reproduire les résultats de l'analyse de l'immunopeptidome de cellules A549-ACE2/TMPRSS2 effectués par Weingarten-Gabbay, S et al. (2021)[62] qui ont gracieusement partagé la lignée cellulaire conçue dans le cadre d'une étude sur l'immunopeptidome de cellules infectées par SRAS-CoV-2. Malheureusement, nous n'avons pas été en mesure de procéder à l'infection des cellules. Nous avons utilisé les résultats des cellules naïves pour le volet de ce travail. Comme ces expériences ont été

analysés à travers deux plate-forme expérimentale et outils d'analyse différents, nous attendons des résultats qui varient de manière significative.

## **4.2 - Matériel et méthodes**

### *Culture cellulaire*

La culture des cellules A549-ACE2/TMPRSS2 a été effectué selon les mêmes conditions décrites à la section 2.2. Les cellules ont été maintenues dans un milieu DMEM/FBS10%/P-S 1% supplémenté de 1µg/mL de puromycin et 1µg/mL de blasticidin comme agents de sélection pour les cellules comprenant les récepteurs transduits. Ces cellules ont été partagés gracieusement par l'université de Boston. Elles sont utilisées car elles ont été transformées pour être susceptible à l'infection par SRAS-CoV2 effectué dans une autre partie de leur expérience.

### *Isolation des peptides du CMH par purification d'immuno-affinité*

L'isolation des peptides a été effectué selon les mêmes conditions décrites à la section 2.2. 100x10<sup>6</sup> cellules ont été utilisés pour chaque expérience.

### *Identification des peptides du CMH de classe I et II par LC-MS/MS*

L'identification des peptides a été effectuée selon les mêmes conditions décrites à la section 2.2. Les résultats de l'acquisition des données ont ensuite été exportés pour être analysés par le logiciel MVP

\*Référer à Weingarten-Gabbay, S et al. (2021) pour la méthode utilisée par le groupe de Boston. Les données brutes des cellules naïves ont été utilisées pour cette expérimentation.

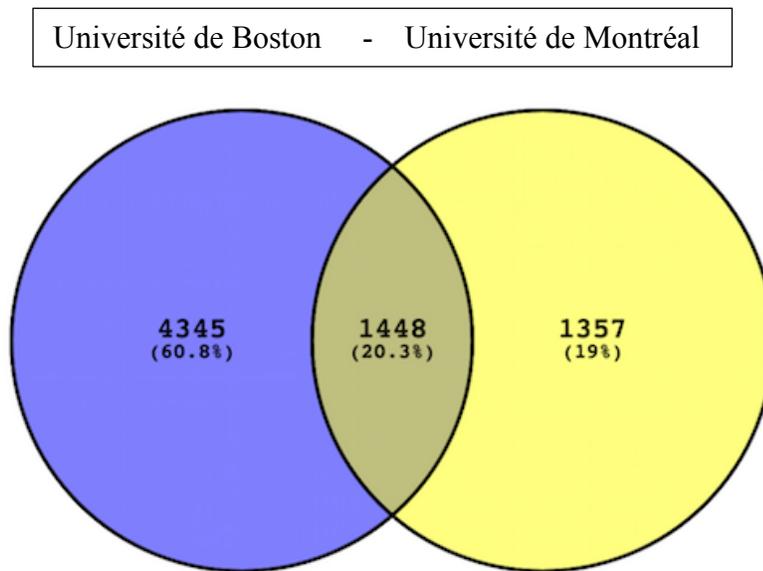
### 4.3 - Résultats

#### 4.3.1 - Reproductibilité entre différents « Pipeline » d'isolation, acquisition et analyse des données inter-établissement.

Dans le but d'investiguer la reproductibilité inter-établissement entre différents protocoles d'isolation, outils d'acquisition et logiciels d'analyse, nous avons commencé par comparer nos résultats avec les résultats de l'analyse d'immunopeptidome de cellules A549-ACE2/TMPRSS2 effectué par un groupe du département de biochimie de l'école de médecine de l'Université de Boston (Weingarten-Gabbay, S et al 2021)[62] qui a gracieusement partagé cette lignée cellulaire avec notre laboratoire. Malgré que cette lignée n'ait finalement pas pu être utilisée pour analyser l'immunopeptidome de cellules infectées avec SRAS-CoV-2 dû à l'impossibilité de recevoir les formations requises à temps, cette lignée a été utilisée comme référence afin d'observer la reproductibilité des méthodes d'isolation et d'analyses entre deux laboratoires utilisant des protocoles d'isolation et des appareils d'analyse différents. Cette lignée cellulaire exprime les allèles HLA A25:01, A30:01, B44:03, B18:01, C12:03 et C16:01. Les données MS/MS du groupe de Boston ont été acquises sur un appareil Orbitrap Exploris 480 (Thermo Fisher Scientific) et interprétées utilisant Spectrum Mill v.7.1 tandis que nos analyses ont été faites sur un Orbitrap Eclipse (Thermo Fisher Scientific) et interprétées à l'aide du logiciel PEAKS X[62].

Comme le montre le diagramme Venn (**Fig. 19**), seulement 20.3 % des peptides ont été identifiés dans les deux expérimentations. Le groupe de Boston a été en mesure d'isoler un nombre considérablement plus élevé de peptides comparativement à nos résultats. De plus, suite à une analyse à l'aide de notre logiciel MVP (**Fig.20**), il est possible d'observer des différences majeures au niveau de la nature des peptides isolés. Dans le cas des cellules analysés par Weingarten-Gabbay, S et al, seulement 6,8% des peptides sont prédit pour se lier fortement aux molécules de l'allèle A30:01 tandis que ce chiffre s'élève à 29.6% des peptides pour ce qui est des analyses faites dans notre laboratoire (**Fig. 20, B**). De manière similaire, 22.2% et 21,8% des peptides sont prédit pour être un ligand fort pour les allèles B18:01 et C12:03 respectivement tandis que nos analyses identifient uniquement 6,1% et 5,4% de ligands forts pour les mêmes allèles respectives

(Fig.20, B). Contrairement aux résultats de Weingarten-Gabbay, S et al qui possède un groupe de 1563 peptides dont 79% sont prédits pour se lier aux molécules de l'allèle B44 :03, très peu de peptides identifiés dans nos analyses sont sujets à interagir avec cet allèle. En fait, seulement 15% d'un groupe de 576 peptides sont prédits comme étant un ligand fort pour l'allèle B44 :03 (Fig. 20, C). De manière surprenante, une plus grande homologie se trouve au sein des allèles HLA-C (Fig. 20, C), un allèle qui est largement sous-étudié dans le domaine de l'immunopeptidome. Cet allèle est moins étudié, notamment dû à sa faible expression sur la surface cellulaire qui représente environ seulement 10% de l'expression des allèles A et B, ce qui complique l'obtention d'échantillons de taille adéquate pour les analyses en SM[152]. L'analyse par MVP nous permet toutefois de confirmer que les peptides isolés par les deux institutions sont en grandes majorité prédits pour s'attacher de manière conforme aux allèles exprimés par cette lignée cellulaire et confirme la spécificité des échantillons.



**Figure 19 :** Diagramme Venn illustrant les peptides communs suite   l'analyse de l'immunopeptidome de 90 millions de cellules A549-ACE2/TMPRSS2 effectu  par deux laboratoires distincts.

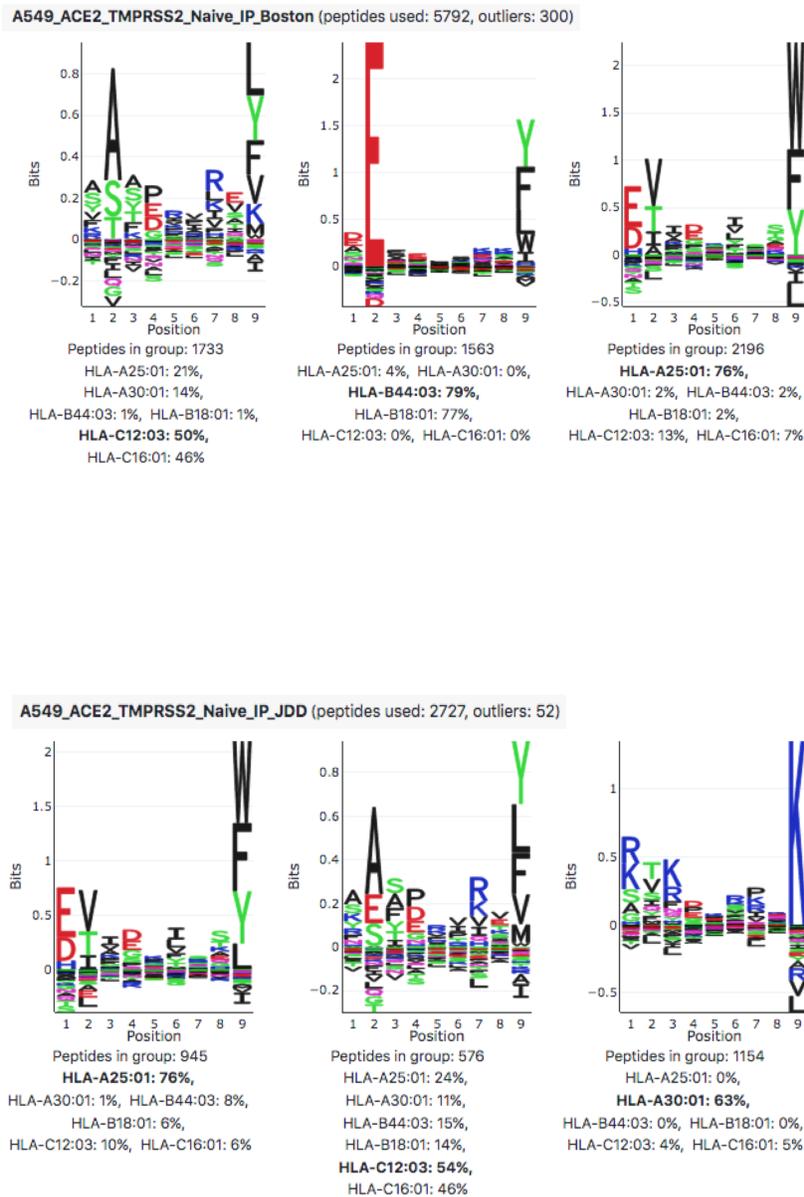
A

Sample	Peptide length	Total peptides	%
A549_ACE2_TMPRSS2_Naive_IP_Boston	all lengths	5792	100
	8-12 mers	5792	100
	other	0	0
A549_ACE2_TMPRSS2_Naive_IP_JDD	all lengths	2805	100
	8-12 mers	2727	97
	other	78	3

B

Allele	Sample	Total peptides	Strong binders	Weak binders	Non-binders
HLA-A*25:01	A549_ACE2_TMPRSS2_Naive_IP_Boston	5792	2098 (36.2%)	1268 (21.9%)	2426 (41.9%)
	A549_ACE2_TMPRSS2_Naive_IP_JDD	2727	864 (31.7%)	344 (12.6%)	1519 (55.7%)
HLA-A*30:01	A549_ACE2_TMPRSS2_Naive_IP_Boston	5792	393 (6.8%)	398 (6.9%)	5001 (86.3%)
	A549_ACE2_TMPRSS2_Naive_IP_JDD	2727	807 (29.6%)	354 (13.0%)	1566 (57.4%)
HLA-B*44:03	A549_ACE2_TMPRSS2_Naive_IP_Boston	5792	1285 (22.2%)	577 (10.0%)	3930 (67.9%)
	A549_ACE2_TMPRSS2_Naive_IP_JDD	2727	166 (6.1%)	169 (6.2%)	2392 (87.7%)
HLA-B*18:01	A549_ACE2_TMPRSS2_Naive_IP_Boston	5792	1265 (21.8%)	896 (15.5%)	3631 (62.7%)
	A549_ACE2_TMPRSS2_Naive_IP_JDD	2727	146 (5.4%)	295 (10.8%)	2286 (83.8%)
HLA-C*12:03	A549_ACE2_TMPRSS2_Naive_IP_Boston	5792	1163 (20.1%)	873 (15.1%)	3756 (64.8%)
	A549_ACE2_TMPRSS2_Naive_IP_JDD	2727	462 (16.9%)	367 (13.5%)	1898 (69.6%)
HLA-C*16:01	A549_ACE2_TMPRSS2_Naive_IP_Boston	5792	969 (16.7%)	762 (13.2%)	4061 (70.1%)
	A549_ACE2_TMPRSS2_Naive_IP_JDD	2727	381 (14.0%)	287 (10.5%)	2059 (75.5%)

C



**Figure 20** – Analyse de l’immunopeptidome de 90 millions de cellules A549-ACE2/TMPRSS2 effectué par deux laboratoires distincts. Université de Boston (A549\_ACE2\_TMPRSS2\_Naive\_IP\_Boston) et Université de Montréal (A549\_ACE2\_Naive\_IP\_JDD). (A) Section « Overview » où l’on peut observer la quantité des peptides qui ont la taille requise pour interagir avec les molécules de CMH-1. (B) Section « Annotation Results », où la prédiction d’affinité des peptides est analysé pour chaque allèle du

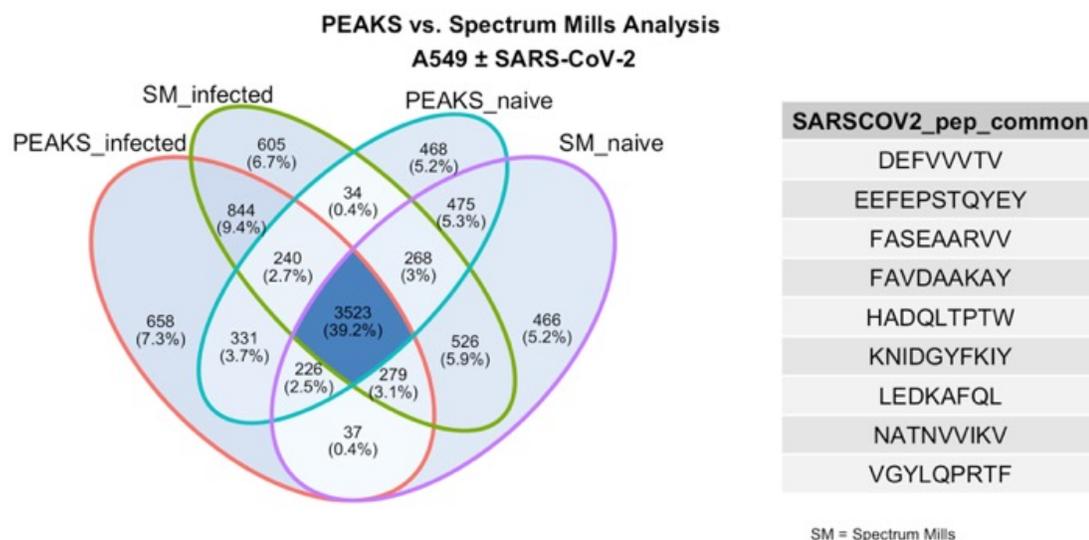
CMH-1 retrouvées dans cette lignée cellulaire. (C) Représentation visuelle de la section « *Sequence Motif* » issue de l'analyse GibbsCluster.

#### 4.3.2 – La disparité dans les logiciels d'analyse de données

Afin d'évaluer si l'analyse de données par différents moteurs de recherche utilisés en protéomique affichent des résultats similaires, nous avons utilisé les données brutes des cellules A549-ACE2/TMPRSS2 avant et après l'infection au coronavirus SRAS-CoV2 de Weingarten-Gabbay, S et al (2021) qui ont été générées par un appareil Orbitrap Exploris 480 (Thermo Fisher Scientific) et traitées avec SpectrumMill v.7.1, pour les analyser avec le logiciel utilisé dans notre laboratoire : PEAKS X. Les résultats ont été frappants.

Comme le montre la figure 21, des différences majeures sont observables entre l'analyse des données brutes par deux moteurs de recherche différents. Au niveau des cellules naïves, 4492 peptides ont été identifiés par les deux moteurs de recherches tandis que 1308 et 1071 ont été exclusivement identifiés par Spectrum Mill et PEAKS respectivement. Au niveau des cellules infectés, 4886 peptides ont été identifiés par les deux moteurs de recherche tandis que 1433 et 1252 peptides uniques ont été identifiés uniquement par Spectrum Mill et PEAKS respectivement (**Fig.21**). De tels taux de variation provenant de l'analyse de données brutes démontrent à quel point il est primordial d'établir des protocoles standardisés et reproductibles au niveau des analyses d'immunopeptidomes et d'assurer l'inter-connectivité des logiciels qui en font le traitement des données.

De manière encore plus importante, le nombre de peptides viraux identifiés par les deux engins de recherche varie aussi considérablement. Les analyses par Spectrum Mill ont permis l'identification de 16 peptides provenant des protéines du protéome de SRAS-CoV-2 tandis que les analyses par PEAKS ont seulement été en mesure d'en identifier 9 qui ont tous aussi été identifiés par les analyses effectués par Spectrum Mill v.7.1. (**Tableau 1A, Fig. 21**).



**Figure 21** – Résultats de l’analyse de données brutes d’immunopeptidome de cellules A549-ACE2/TMPRSS2 avant et après infection par SRAS-CoV2 analysés par deux différents moteurs de recherches. SM\_naive : cellules naïves analysés par Spectrum Mill v.7.1, PEAKS\_naive : cellules naïves analysés par PEAKS X. SM\_infected : cellules infectées par SRAS-CoV2 analysés par Spectrum Mill v.7.1, PEAKS\_infected : cellules infectées par SRAS-CoV2 analysés par PEAKS X, SARSCOV2\_pep\_common : peptides viraux identifiés dans les deux expérimentations.

#### 4.4 - Discussion : L’harmonisation des données inter-institution est déficiente dans le domaine de l’immunopeptidomique.

Comme constaté dans la comparaison des résultats provenant de l’analyse d’une même lignée cellulaire analysé par deux institutions ainsi que les résultats divergents lorsque des données brutes de spectrométrie de masse ont été analysées par deux moteurs de recherche distincts, des différences majeures sont observées entre les résultats. L’hétérogénéité des résultats peut être causée par de nombreux facteurs. Non-seulement les analyses ont été faites sur une instrumentation différente et interprétées avec des logiciels différents mais les méthodes d’isolation des peptides diffèrent au niveau des tampons de lyse et d’élution utilisés[62].

Les divergences notables entre les résultats d’analyse provenant d’une même lignée cellulaire entre deux établissements démontrent le chemin restant à parcourir afin d’intégrer l’étude de l’immunopeptidomique au même titre que la génomique et

transcriptomique au niveau de l'étude de la biologie des systèmes et ce, par l'utilisation de techniques de contrôle de la qualité des échantillons et par reproductibilité des protocoles et de la performance des appareils d'analyse.

Il est à noter que les expériences effectuées dans le cadre de ce travail ont été faites par un utilisateur ayant une expérience très limitée avec les protocoles d'isolation de peptides. Les résultats de ces expériences sont le résultat des premières tentatives d'isolation des peptides du CMH par l'expérimentateur. Il est possible qu'avec plus d'expérience, un nombre plus élevé de peptides pourrait être isolé, surtout pour les lignées cellulaires adhérentes qui sont plus contraignantes au niveau de la récolte.

L'identification des peptides traités par la machinerie métabolique qui parcourent les voies de présentation de l'antigène pour se retrouver à la surface des molécules du CMH-1 est une étape cruciale dans le développement d'immunothérapies et de vaccins basés sur la réponse T[61, 95]. Il est donc important d'assurer le développement de contrôle de la qualité au niveau de l'échantillonnage et d'homogénéiser le traitement des données par les différents moteurs de recherches afin d'atteindre une reproductibilité dans les résultats de telles analyses.

#### **4.5 - Modèle d'analyse d'immunopeptidome de cellules susceptibles aux coronavirus : MRC-5 et OC43**

Depuis le début de la pandémie mondiale créée par l'apparition du virus SRAS-CoV2, un effort concerté entre divers gouvernements et compagnies pharmaceutiques a permis le développement de plusieurs vaccins basés sur diverses plateformes, et ce, en un temps record. Ces différents vaccins conçus jusqu'à maintenant comprennent en grande majorité la séquence de l'ARNm du domaine d'attache des récepteurs de la protéine Spike et ont été efficace pour monter une réponse immune en produisant des anticorps contre cette protéine de l'enveloppe externe du virus[153]. Toutefois, de plus en plus d'évidences démontrent que cette protéine externe est sujette à plus de mutations que certaines protéines virales qui se retrouvent dans les parties internes du virus[62, 130, 154].

Les anticorps résultant de l'immunisation à la protéine « spike » ont l'avantage de s'associer aux protéines responsables de l'attachement viral et permettent ainsi de prévenir l'infection par le virus. Le problème inhérent de cette approche sont les titres d'anticorps qui doivent rester très élevés pour être efficaces, ce qui n'est pas nécessairement une approche viable pour un vaccin qui a besoin d'être efficace à long terme. Cette approche cause donc problème dû aux mutation inévitables du virus qui pourraient ultimement mener à contourner la protection offerte par le vaccin et à la possibilité d'observer une perte de la capacité de neutralisation plusieurs mois après l'immunisation dû à un titre d'anticorps insuffisant pour contrôler l'infection[61, 92].

Pour les raisons précédentes, plusieurs groupes se penchent sur le développement d'un vaccin qui produirait une réponse des cellules T envers de multiples protéines des structures externes et internes du virus, plus précisément, des sections du génome qui ont moins de chances de muter. Malgré que les cellules T n'empêchent pas l'infection, principalement dû au fait qu'ils s'activent uniquement suite à l'apparition de celle-ci et à l'identification d'une cellule infectée, elles restent des acteurs importants dans le contrôle et dans l'élimination d'un pathogène dans l'organisme, un tel vaccin serait une alternative plus sécuritaire pour les patients à risque[61, 95]. De tels vaccins pourraient donc avoir comme résultats de plus faibles symptômes et une réduction de la transmission chez les personnes infectées. De plus, des études ont montré que les personnes qui ont été infectées par SRAS-CoV-2 génèrent généralement des lymphocytes T qui ciblent au moins 15 à 20 fragments différents de protéines du coronavirus[130]. Étant donné que les fragments de protéines présentés sur les molécules du CMH peuvent varier considérablement d'une personne à l'autre, une population générera une grande variété de lymphocytes T susceptibles d'empêcher la propagation du virus par l'entremise d'une immunité de groupe soutenue par un répertoire élargie de lymphocytes T au sein d'une communauté. Il est donc très difficile pour un virus de muter à un rythme assez rapide pour échapper à la reconnaissance cellulaire à multiples facettes offerte par la variété des fragments utilisés pour la reconnaissance du virus par les cellules T, situation contraire à ce qui se déroule présentement pour la réponse aux anticorps produits par la protéine

« spike » qui semble démontrer une baisse de potentiel de neutralisation pour certains variants [130, 154, 155].

Une des limitations majeures avec l'analyse de cellules infectées par SRAS-CoV2 est la nécessité de travailler sous des conditions de biosécurité de niveau 3. Comme la majorité des institutions n'ont pas accès à ce genre de facilités, et dû aux nombreuses formations et contraintes nécessaires afin de manipuler des pathogènes dangereux, nous avons analysé la possibilité d'utiliser un modèle de Coronavirus circulant de manière endémique chez l'humain qui est responsable de faibles symptômes d'apparence grippales ne représentant pas de risques graves au niveau des manipulations : le virus HCoV-OC43 [156]. La lignée cellulaire MRC-5 a été utilisée comme modèle cellulaire. Cette lignée est utilisée car elle est permissive au virus sans à avoir à y effectuer des modifications et elle provient de cellules de poumon, organe ciblé par le virus HCoV-OC43. La pertinence de l'étude des membres des *Sarbecovirus*, le sous-génome viral contenant SRAS-CoV-1 et 2, a été établie par l'identification d'anticorps et de cellules T avec potentiel de neutralisation croisée pour plusieurs  $\beta$ -Coronavirus ainsi que par la découverte de cellules T réactives à SRAS-CoV2 chez des individus non-exposés au Covid-19[92, 126, 157]. En effet, une infection par un coronavirus circulant chez l'humain peut permettre la présentation d'un épitope viral dont la séquence peut être partagée par de nombreux coronavirus[92, 158]. L'identification de ces régions communes à de multiples coronavirus, en mettant l'accent sur les séquences ayant les plus faibles chances de subir des mutations, serait une étape importante dans le développement de vaccins de seconde génération, non seulement dans le but d'éradiquer le SRAS-CoV2, mais aussi dans le but de préparer le terrain dans la conception de vaccins efficaces contre d'autres coronavirus circulant chez des animaux qui représentent un potentiel danger pour s'adapter à l'espèce humaine dans de futures générations[126].

Le but de cette section du travail (objectif 5) est d'appliquer les méthodes et outils de contrôle de qualité des objectifs 1, 2 et 3 pour analyser l'immunopeptidome de cellules MRC-5 avant et après infection avec Coronavirus HCoV-OC43 qui circule chez l'humain.

## 4.6 – Matériel et méthodes

**Note 1 :** Compte tenu que les manipulations sont effectuées sur des cellules infectés, les échantillons destinés à être analysés en cytométrie de flux ou par spectrométrie de masse ont été inactivés par une solution de Paraformaldéhyde (PFA) 4% pour une durée de 30 minutes.

**Note 2 :** Toutes les étapes des protocoles de quantification du CMH et de l'immunoprécipitation ont été effectués sous une hotte biologique dans une salle à niveau de confinement 2.

### *Culture cellulaire.*

Les cellules MRC-5 ont été maintenues dans un milieu EMEM, FBS 10%, Acides aminés non-essentiels 1%, NaPO<sub>4</sub> 1% et pénicilline/streptomycine 1%. 10 pétris de 10x245mm ont été utilisés, 5 pour les cellules naïves et 5 pour les cellules infectées.

### *Infection*

Après 4 jours d'incubation des cellules MRC5 et l'obtention d'une confluence de 80% dans la totalité des pétris, 100µL de l'extrait viral qui nous a été donné par le laboratoire du docteur Roger Lippé au CHU St-Justine (OC43 Exp 330 P3 Media Fraction VR-1558 beta-coronavirus OC43) est ajouté à 60mL de milieu RPMI/0.1% BSA. 60mL de milieu RPMI/0.1% BSA sans virus (*Mock*) est préparé pour être utilisé avec les cellules non infectés. Le milieu de culture EMEM est retiré des 10 plaques 10x245mm par aspiration manuelle. 12 mL du milieu d'infection contenant le virus ou sans le virus (*Mock*) sont ajoutés par plaque. L'adsorption est effectuée pendant 1h dans un incubateur à 37C°, 5% CO<sub>2</sub>. 48 mL de milieu réduit (EMEM 2% BGS 1% NEAA 1% NaPO<sub>4</sub> 1% P/S) sont ajoutés pour chaque plaque pour un total de 60 mL de milieu de culture par plaque. Les plaques sont incubées comme précédemment à 37C°, 5% CO<sub>2</sub> jusqu'à ce que les effets cytopathiques soient visibles et que l'infection soit avancée, définie ici comme >20%, <50% de cellules infectées et en suspension (desquamation) afin préserver un nombre

suffisant de cellules vivantes pour les quantités requises en SM. Lors de cette expérimentation, les cellules ont été recueillies 4 jours après l'infection. Pour la récolte, les cellules sont ensuite détachées par trypsination (1 min à 37C° d'incubation avec 3 mL de solution de trypsine à 0,25% par pétri) et inactivés par une solution de Paraformaldéhyde (PFA) 4% pour une durée de 30 minutes.

#### *Analyse quantitative du MHC-1 par immunofluorescence*

Les mêmes étapes observées en section 2.2 ont été utilisées pour l'analyse de l'abondance des molécules du CMH-1 de surface des cellules MRC-5 infectées et naïves. Les cellules ont été assujetties à un traitement de 60 minutes dans une solution de PFA 4% afin d'inactiver les particules virales pouvant se retrouver dans les échantillons.

#### *Isolation des peptides du CMH des cellules MRC-5 par purification d'immuno-affinité*

Le protocole décrit à la section 2.2 a été observé pour l'isolation des peptides du CMH des lignées MRC-5 naïves et infectées. Toutefois, toutes les étapes ont été effectuées sous la protection d'une enceinte biologique dans une salle de niveau de confinement 2.

33 millions de cellules ont été utilisées

#### *Identification des peptides du CMH de classe I et II par LC-MS/MS*

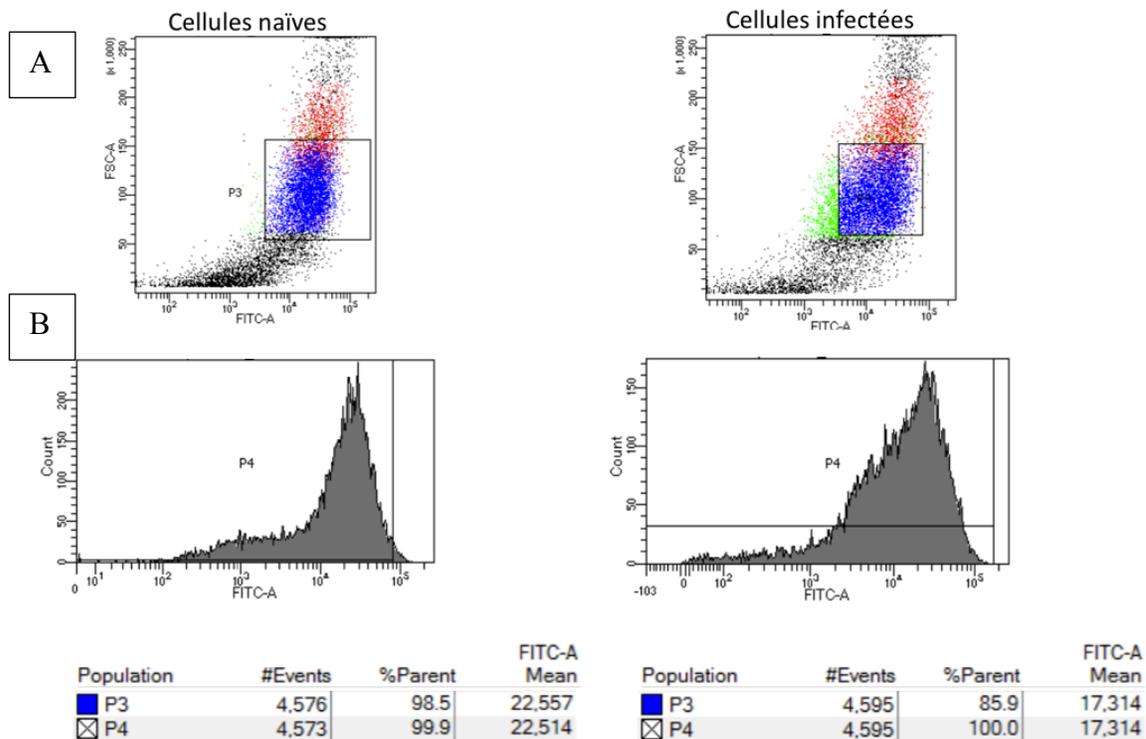
33 millions de cellules naïves et infectées ont été analysées. L'identification des peptides a été effectuée selon les mêmes conditions décrites à la section 2.2. Les résultats de l'acquisition des données ont ensuite été exportés pour être analysés par le logiciel MVP.

## **4.7 – Résultats**

### **4.7.1 – Les cellules MRC-5 affichent une forte expression des molécules du CMH-1**

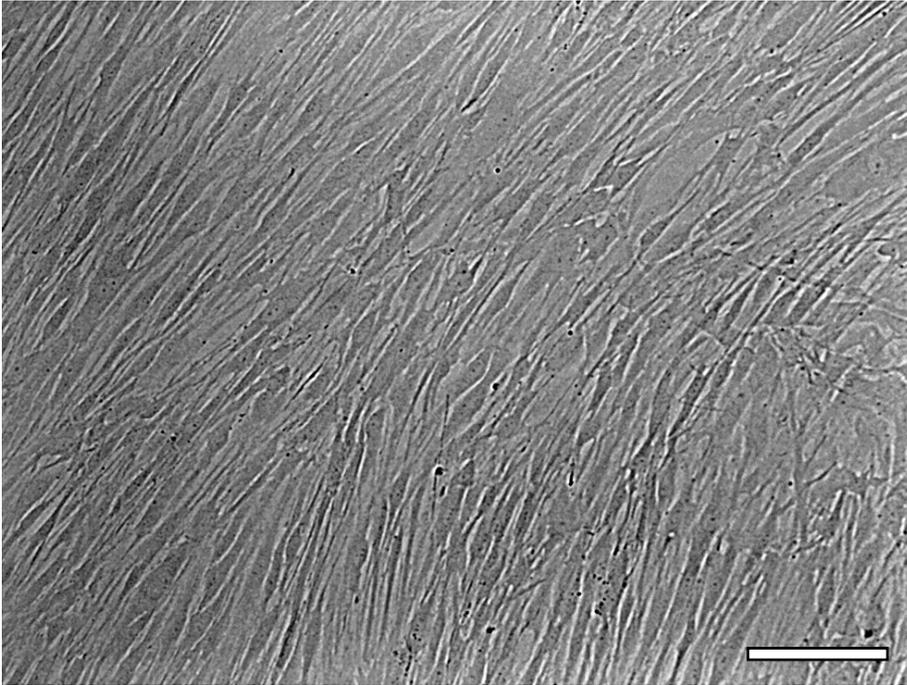
Étant donné qu'un facteur limitant à une analyse appropriée des ligands du CMH nécessite l'utilisation d'une lignée cellulaire exprimant de fort taux des molécules HLA,

nous avons analysé l'abondance de surface des récepteurs du CMH-1 sur les cellules MRC-5. Les résultats démontrent une forte expression de ce marqueur de surface. En termes de comparaison, elles expriment 66,59% des molécules exprimés par notre lignée JY utilisé comme contrôle de la qualité pour les cellules humaines. De plus, malgré une baisse 23.25% de l'expression des molécules du CMH suite à l'infection par HCoV-OC43, les cellules conservent tout de même une forte expression relative de récepteurs du complexe majeur d'histocompatibilité à leur surface (**Fig.22**). Toutefois, une faible proportion des cellules semble avoir perdue toute expression des molécules du CMH-1(12,6%), comme quoi la présence du virus affecte probablement la migration de ces complexes à la surface de la membrane cellulaire. Ces cellules, utilisées conjointement au virus OC-43, sont donc considérées comme de bons candidats pour l'analyse de l'immunopeptidome de cellules infectées par un membre de la famille des coronavirus.

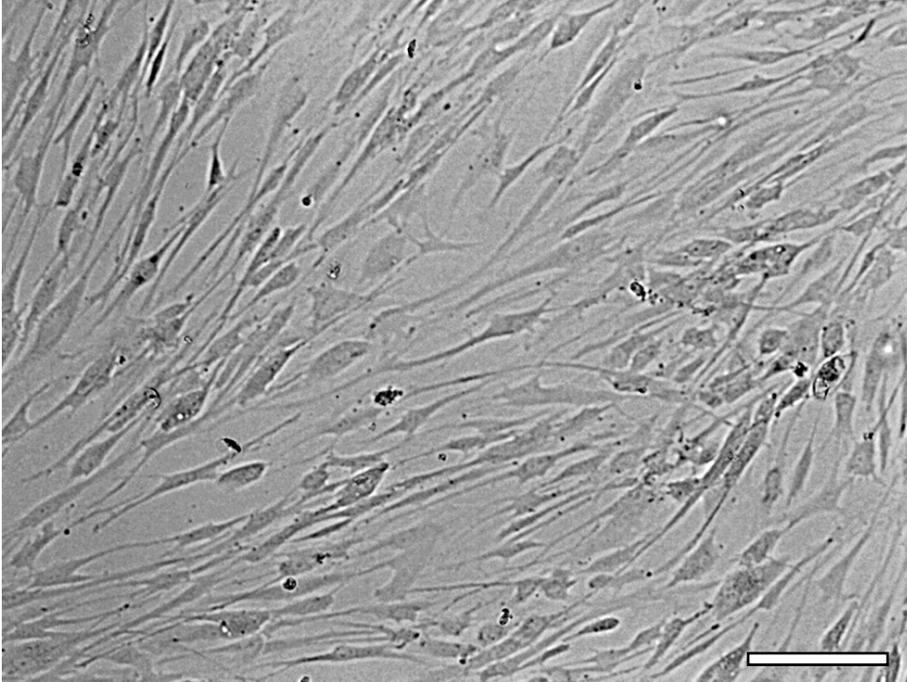


**Figure 22** - Résultats de cytométrie de flux de l'expression des molécules du CMH-1 sur des cellules MRC5 avant et après l'infection par coronavirus OC43. (A) Fenêtre d'acquisition des cellules dont les molécules du CMH-1 sont marquées par fluorescence. (B) Intensité de la fluorescence des cellules sélectionnées en A.

A)



B)



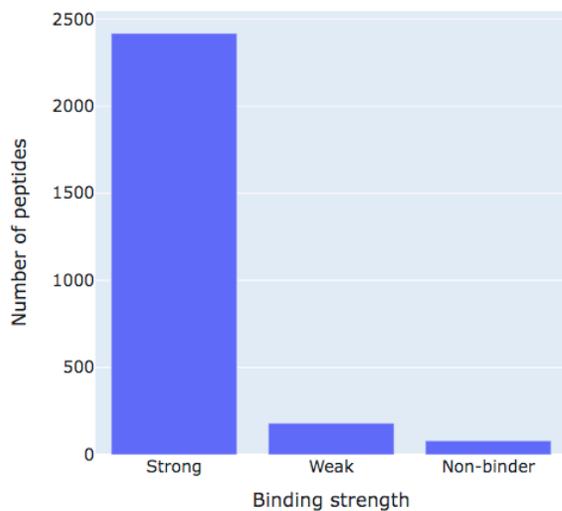
**Figure 23** – Images microscopiques avec un grossissement de 40X. Barre d'échelle : 100um

A) Cellules MRC-5 naïves. B) Cellules MRC-5 infectés par le virus HCoV-OC43 pour une période de 3 jours

#### 4.7.2 – L'immunopeptidome des cellules MRC5 naïves

Comme prévu par les analyses préalables au niveau de la quantification des molécules du CMH-1 de surface, les analyses de spectrométrie de masse faites sur les échantillons de peptides isolés ont révélé un ensemble complexe de segments de protéines composant l'immunopeptidome de cellules MRC5. En effet, 2678 peptides de taille comprise entre 8 et 12 acides aminés ont été identifiés par SM (**Ann. 5**). Les peptides qui composent cet ensemble proviennent de protéines ayant diverses fonctions, que ce soit structurelles : P12111-CO6A3, composante de la matrice extracellulaire, acteurs dans les voies de signalisation : P35222-CTNB1, composante clef de la voie Wnt ou même ayant des fonctions catalytiques : Q10570-CPSF1, sous-unité du complexe de clivage et polyadénylation.

Grace aux analyses faites par MHCvizPipe, il est possible de faire la première étape du contrôle de la qualité qui est de s'assurer que les peptides identifiés sont spécifiques aux allèles du CMH exprimés par les cellules utilisées dans le modèle d'étude. Dans ce cas-ci, les cellules MRC5 expriment les allèles HLA suivantes : HLA-A02:01, A29:02, B07:02, B44:02, C05:01 et C07:02. Comme démontré en figure 24, parmi les 2678 peptides, 2418 peptides sont identifiés pour interagir de manière forte avec un ou plusieurs des allèles du CMH exprimés par les cellules, s'assurant ainsi la vérification du critère important de la spécificité des échantillons analysés. De plus, si l'on tient compte des ligands faibles, 97 % des peptides composant l'échantillon interagit avec les molécules du CMH. Ce qui confirme que les techniques d'isolation des peptides sont très spécifiques pour les peptides du CMH et entraînent très peu de contaminations.



**Figure 24** – Diagramme à barres illustrant la proportion des peptides isolés dans l'échantillon de cellules MRC5 naïf ayant une forte (strong), faible (weak) ou sans (non-binder) affinité pour les allèles du CMH exprimés par cette lignée cellulaire.

#### 4.7.3 – Immunopectidome des cellules MRC5 infectées par virus OC43.

L'infection par ce virus à un profond impact sur la nature et l'abondance des peptides qui compose l'immunopectidome des cellules infectées. Malheureusement, l'échantillon de cellules infectés était considérablement plus petit que prévu. En effet, les cellules ont été infectées trop longtemps avant le moment de la récolte (4 jours) avec une confluence d'environ 20 % par rapport au cellules naïves qui affichaient une confluence de 100 %. L'échantillon non-infecté totalisait 116 millions (33 millions ont été utilisés pour l'analyse) de cellules tandis que l'échantillon infecté totalisait 33 millions de cellules. Pour des analyses subséquentes, un temps d'infection étalé sur 3 jours produirait une meilleure récolte de cellules.

L'échantillon de cellules infectées était composé par seulement 92 peptides qui rentrent dans les critères de taille entre 8 et 12 acides aminés, 86 peptides provenant de peptides humains et 6 (**Ann. 5**) provenant des protéines P36334, P30215 et Q80872 du génome viral. Encore une fois, les analyses faites par MVP confirment la spécificité des peptides des échantillons pour les allèles HLA exprimés par les cultures cellulaires. En

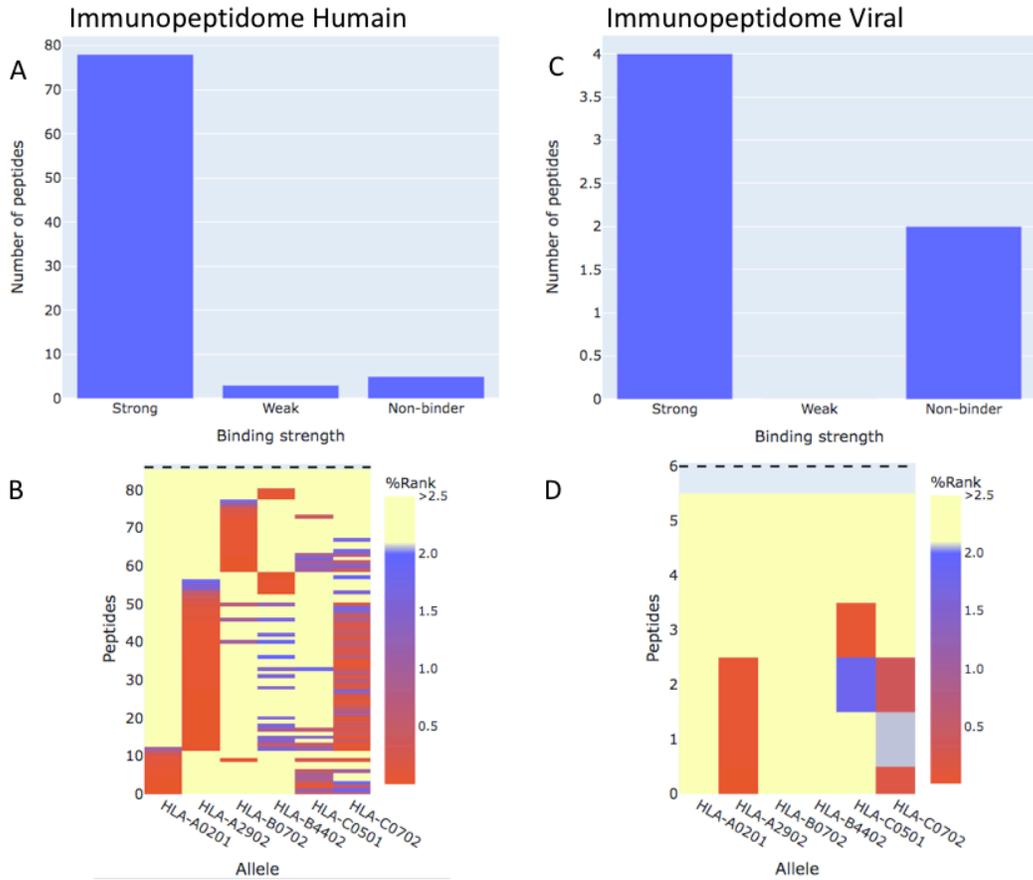
effet, 94% des peptides humains et 67% des peptides viraux sont prédits pour interagir avec les allèles HLA des molécules du CMH exprimés par les cellules MRC5 (**Fig. 25**). De manière intéressante, le peptide viral de 10 acides aminés SGFAVYVKS se retrouve aussi flanqué des résidus V et G dans sa portion C-terminale pour former le peptide SGFAVYVKS<sub>10</sub>VG. De plus, dans les 9 peptides viraux isolés, 3 peptides ont été exclus de par leur taille : TSGFAVYVKS<sub>10</sub>VG, DTSGFAVYVKS<sub>10</sub>VG et TSGFAVYVKS<sub>10</sub>VGNYR. Ces peptides, malgré que leur taille soit non-conventionnelle pour interagir avec les molécules du CMH-1, possèdent toutes le même noyau composé de la séquence SGFAVYVKS auquel est flanqué les résidus T ou DT en position N-terminale et NYR en C terminale. Il n'est pas impossible pour des peptides de plus longue taille de se retrouver sur les molécules du CMH-1 par le phénomène de renflement. Toutefois, lorsque ce phénomène se produit, ce sont généralement les acides aminés situés en positions d'ancrages qui sont conservés[159]. Néanmoins, il est possible que ces peptides aient entrepris des chemins de dégradation alternatifs résultant en la formation de peptides de longueurs différentes et réussissent à se loger dans la niche peptidique en gonflant leur section médiane lors de leur chargement sur les molécules du CMH-1. Ces peptides proviennent de la protéine source Q01455 qui est codé par le gène M du virus. Ce gène est responsable de coder pour l'enveloppe virale et cette protéine joue un rôle central dans la morphogénèse et l'assemblage du virus via ses interactions avec d'autres protéines virales (<https://www.uniprot.org/uniprot/Q01455>). Les autres peptides identifiés proviennent des protéines sources P36334, P30215 et Q80872. La glycoprotéine P36334, ou protéine « *Spike* » est codée à partir du gène S, qui a des fonctions d'attachement à la cellule ainsi qu'à la fusion entre les membranes virales et cellulaires (<https://www.uniprot.org/uniprot/P36334>). P30215 provient du gène HE et à une fonction structurelle en formant de courtes pointes à la surface du virus qui potentiellement, servent de protéine secondaire d'attachement viral pour initier l'infection, la protéine « *spike* » étant la principale. Cette protéine contribue aussi à empêcher l'auto-agrégation et assure une propagation efficace de la progéniture du virus (<https://www.uniprot.org/uniprot/P30215>). Finalement, Q80872 est une protéine non-

structurelle dont les fonctions sont très peu caractérisées dans la littérature (<https://www.uniprot.org/uniprot/Q80872>).

On peut observer les scores d'affinité pour le CMH-1 attribués pour chaque peptide avec les différents allèles retrouvés dans le modèle cellulaire au tableau 2 (page 104). Aucun des peptides n'interagit avec l'allèle A\*02 :01, B\*07 :02 et B\*44 :02 tandis que les peptides FINSMQFNY, NINPHNYSY et HFSYVPTKY interagissent de manière forte avec l'allèle A\*29:02. TADEAIKF interagit de manière forte et NINPHNYSY de manière faible avec les molécules de l'allèle C\*05 :01 et finalement, NINPHNYSY et HFSYVPTKY interagissent de manière forte avec les molécules C\*07 :02.

L'immunopeptidome viral identifié par SM provient donc d'une quantité limitée de protéines mais il est intéressant de constater que la source de ces protéines varie grandement et que la machinerie cellulaire est en mesure de dégrader non-seulement les protéines externes du virus mais aussi des protéines internes ou structurelles qui théoriquement, sont moins soumises aux pressions de sélection et ont moins tendance à muter. Les allèles les plus propices de porter ces peptides sont les allèles A\*02 :01 et C\*07 :02

- Les immunopeptidomes complet des cellules MRC5 naïves et infectés ont été ajoutés en Annexe 5.



**Figure 25** – Diagrammes à barres et « *Heat-Maps* » illustrant la proportion des peptides isolés dans les échantillons de cellules MRC5 naïves (A, B) et infectées (C, D) ayant une forte (strong), faible (weak) ou sans (non-binder) affinité pour les allèles du CMH exprimés par cette lignée cellulaire ainsi que la distribution des peptides pour chaque allèle exprimé.

**Tableau 2** - Rang percentile des prédictions d'affinités des peptides viraux générés par NETMHCPan avec les allèles correspondants. (Vert : forte affinité, Rouge : faible affinité, Noir : sans affinité) (INCLUDE PROTEINE SOURCE QUI SE TROUVE A LA PAGE 144)

Peptide	Longueur	A0201	A2902	B0702	B4402	C0501	C0702	Source
TADEAIKF	8	16.60	15.530	6.9070	8.4904	0.0600	3.2223	Q01455
FINSMQFNY	9	9.9103	0.0110	21.8571	11.2436	4.9150	2.0612	Q80872
NINPHNYSY	9	7.2464	0.0323	4.9072	2.4181	1.8279	0.3620	P30215
HFSYVPTKY	9	15.5689	0.0050	13.6651	2.6913	4.4052	0.2098	P36334
SGFAVYVSKK	10	48.0	19.3958	47.750	40.50	54.0	32.2	Q01455
SGFAVYVSKKVG	12	87.50	90.00	90.0	95.0	90.0	100.0	Q01455
TSGFAVYVSKKVG	13	95.00	100.0	95.0	100.0	100.0	100.0	Q01455
DTSGFAVYVSKKVG	14	100.0	95.0	100.0	100.0	100.0	100.0	Q01455
TSGFAVYVSKKVGNYR	16	87.50	64.0	85.0	80.0	80.0	85.0	Q01455

#### 4.8 – Discussion : Les avantages et inconvénients de l'utilisation du modèle MRC-5/OC43

Comme conclu au paragraphe précédent, un des principaux atouts des cellules MRC-5 utilisées avec le virus HCoV-OC43 est leur forte expression des molécules du CMH-1 qui est suffisamment conservée lors de l'infection. De plus, les contraintes liées à l'utilisation d'un tel modèle d'étude sont considérablement simplifiées au niveau organisationnel et bureaucratique contrairement à lorsqu'il est question de manipulations et expérimentations avec des pathogènes dangereux tels que SRAS-CoV2. Les salles de niveau de confinement de niveau 3 sont non-seulement très rares au niveau des institutions de recherche, mais de plus, les formations pour y accéder sont rigoureuses, s'étalent sur de longue durée et sont dispensées par un personnel hautement spécialisé dont les disponibilités peuvent être restreintes et varient selon les institutions. Il est donc nettement plus facile pour un utilisateur moyennement expérimenté de manipuler et de procéder à l'infection de cellules nécessitant des mesures de confinement moins contraignantes.

L'utilisation de ces cellules comporte toutefois quelques inconvénients, la technique la plus répandue pour déterminer le titre d'un échantillon viral est test de

plaque (*Plaque Assay*). Contrairement aux modèles couramment utilisés pour la titration d'échantillon de SRAS-CoV2, le virus HCoV-OC43 ne provoque pas de plages de lyses bien définies, ni sur les cellules MRC-5, ni sur des cellules VERO E6 couramment utilisé dans ce type de procédures. Il est donc impossible d'utiliser cette méthode facile et peu coûteuse pour la quantification du virus. À noter aussi que la cinétique de l'infection est considérablement plus lente chez le virus OC43. La figure 23 illustre les effets cytopathiques du virus OC-43 sur les cellules modèles. On y observe non seulement des changements dans la morphologie des cellules, mais aussi un déclin considérable dans le nombre de cellules. Un tapis de cellules confluentes à environ 80 % avait été infecté pour une période de 3 jours avant la prise des images.

En plus des avantages au niveau de l'expression du récepteur d'intérêt, comme le SRAS-CoV2, le virus HCoV-OC43 est un  $\beta$ -coronavirus qui attaque les voies respiratoires et qui est également transmis par voie aérienne, faisant de lui un modèle de recherche homologue sur multiples facteurs[160].

## CHAPITRE 5 – Discussion et perspectives

### 5.1 - L'avenir de l'étude du complexe majeur d'histocompatibilité au sein de l'approche de la « biologie des systèmes »

De nombreuses avancées dans les technologies de chimie analytique ont permis des percées majeures dans le domaine de l'oncologie. L'analyse par spectrométrie de masse des peptides associés aux molécules HLA est devenue un nouveau sujet d'intérêt en immunologie et dans le développement de vaccins et immunothérapies. La capacité actuelle des technologies de spectrométrie de masse à identifier des milliers de peptides associés au CMH à partir de tissus humains, y compris des néo-antigènes spécifiques à certaines tumeurs, suscite un grand intérêt pour la médecine de précision[96]. La médecine courante est largement concentrée sur le traitement des maladies suite à l'apparition des premières manifestations de celles-ci mais la médecine moderne passe progressivement du traitement à la prévention des maladies en fonction du risque individuel. Des études à grande échelle dans le domaine de la génomique ont permis l'identification de plusieurs variables génétiques qui prédisent la prédisposition d'un individu à certaines maladies courantes. En fait, l'identification robuste, rapide et peu coûteuse de variables génétiques chez les individus rend aujourd'hui possible des approches de médecine qui permettent d'identifier la prédisposition d'un individu à une maladie et permet de lui offrir des traitements personnalisés[60, 96, 161].

Les régions qui codent pour les gènes HLA ont été établies comme les régions du génome associées au plus grand nombre de maladies chez l'humain, dont la sclérose en plaques, le diabète de type 1, l'arthrite rhumatoïde, le lupus et le lymphome de Hodgkin[162, 163]. Des centaines de loci dans la région HLA rendent susceptibles et prédisposent certains individus à plusieurs maladies auto-immunes en développant des auto-anticorps qui s'attaquent à divers tissus de l'hôte, notamment la gaine de myéline des axones dans le cas de la sclérose en plaque[163]. La plupart des variations des gènes HLA sont associées à des maladies du système immunitaire, mais les mécanismes exacts à l'origine de ces associations restent largement méconnus [60].

Des preuves récentes indiquent que l'analyse de l'immunopeptidome de milliers d'individus peut révéler un vaste éventail de signatures immunopeptidomiques ou la présence de biomarqueurs qui pourraient être utilisés dans un contexte clinique pour prédire la susceptibilité ou la résistance d'un individu aux maladies immunitaires ainsi que sa prédisposition à répondre aux vaccins et aux immunothérapies[60, 110, 164]. Dans un avenir à moyen terme, la quantification et caractérisation des immunopeptidomes à grande échelle deviendra une réalité. La réalisation de cet objectif nécessitera un développement technologique considérable allant des techniques hautement standardisées pour l'isolement à haut débit des peptides jusqu'aux ressources informatiques appropriées pour traiter l'analyse de très grands ensembles de données immunopeptidomiques générés par ceux-ci. Néanmoins, une fois que cette science aura rattrapé son retard comparativement aux analyses génomique, protéomique et transcriptomique, l'immunopeptidomique aura une grande incidence sur la gestion personnalisée des soins de santé et sur les soins de santé publique.

Malheureusement, des limitations techniques persistantes ont jusqu'à maintenant empêché le déploiement et l'utilisation des outils d'analyses de l'immunopeptidomiques en immunologie. L'analyse des immunopeptidomes à l'aide des technologies de spectrométrie de masse est un processus relativement complexe qui nécessite trois composantes principales : l'isolement des peptides associés aux HLA par immuno-précipitation par affinité, l'acquisition des peptides isolés à l'aide d'une approche de spectrométrie de masse et l'analyse informatique des peptides acquis à l'aide d'outils logiciels spécialisés. À présent, les principaux défis dans le domaine de l'immunopeptidomique sont liés à ces différentes étapes. Il existe une incertitude quant au rendement des procédures d'isolement des peptides, de plus, ce type d'analyse requiert une quantité de matériel biologique considérable. Le coût, la sensibilité et la reproductibilité des technologies de spectrométrie de masse entre les divers instruments sont aussi des barrières qui empêchent l'intégration de cette science au même titre que la génomique, transcriptomique ou protéomique. À ceci s'ajoute le manque de normes expérimentales et informatiques de haute qualité et la faible accessibilité au grand

volume de données de spectrométrie de masse générées par la communauté. Par conséquent, le domaine se développe lentement et très peu de laboratoires dans le monde ont développé l'expertise technique nécessaire pour générer des ensembles de données immuno-peptidomiques de très haute qualité. Tous les éléments mentionnés démontrent qu'il est primordial d'établir des méthodes de contrôle de la qualité et de la spécificité des échantillons analysés dans le but d'atteindre une harmonisation des données[60, 96, 101].

La « biologie des systèmes » (*Systems Biology*) intègre des approches techniques, physiques et mathématiques et les lie à des connaissances biologiques et médicales dans un processus qui permet de visualiser les événements interconnectés au sein d'une cellule. Un effort multidisciplinaire et l'intégration de l'étude de l'immuno-peptidomique est nécessaire pour convertir les informations contenues dans les ensembles de données multidimensionnelles afin de permettre l'identification de biomarqueurs permettant de classer les pathologies en fonction de leur pronostic et de leur réponse aux traitements dans le but d'identifier les cibles optimales dans le cadre de thérapies. Dans un avenir rapproché, les approches de biologie systémiques vont inclure l'étude de l'immuno-peptidomique dans le but d'identifier des biomarqueurs qui augmenteront le taux de réussite et accéléreront la mise en œuvre de thérapies personnalisées efficaces dans l'arsenal clinique des traitements d'immuno-thérapie[96, 98, 165].

### **5.1.2 - Le rôle du CMH dans le cadre des cancers, maladies auto-immunes et des maladies infectieuses.**

Depuis l'apparition des premières formes du système immunitaire, une course aux armements se déroule entre les organismes possédant ces systèmes et les pathogènes qui tentent d'y élire domicile. Chez l'humain, les infections par différentes bactéries, virus et parasites ont contribué de manière significative à la diversification des gènes HLA qui sont reconnus pour être les gènes les plus polymorphiques au sein de la population[60]. D'un côté, cette diversification assure à une population la capacité de certains individus à présenter des segments protéiques de ces pathogènes à la surface de leurs molécules du

CMH et engendre l'effet de l'immunité de masse contre une infection[162]. Le revers de la médaille de cette grande diversité au sein de ce groupe de gènes est l'apparition de nombreuses maladies liés au polymorphisme de certains nucléotides[162]. Ces maladies génétiques affectent parfois le système immunitaire mais elles provoquent aussi plusieurs pathologies qui n'ont aucun lien avec la défense immunitaire [166]. Dans le cas de la rhinite allergique, la présentation clinique la plus courante de l'allergie, une méta-analyse récente a identifié un polymorphisme au niveau d'acides aminés situés dans les régions qui codent pour le sillon de liaison des peptides des molécules du CMH des allèles HLA-DQB1 et HLA-B[60]. Pour le cancer du cancer du poumon, l'analyse de la variation génétique du système HLA a aussi révélé la présence de variantes d'acides aminés dans le sillon de liaison au peptide du système HLA-B\*08:01[60]. Également, dans le cas de l'alopécie, l'une des maladies auto-immunes les plus répandues, une autre analyse a notamment mis en évidence le système HLA-DR comme principal facteur de risque de la maladie avec des variations au niveau des acides aminés situés dans le sillon de liaison au peptide[60]. Au niveau des maladies infectieuses, une analyse incluant plus de 200 000 individus d'ascendance européenne a suggéré un impact important du polymorphismes spécifiques d'acides aminés dans le sillon de liaison des peptides pour huit maladies infectieuses courantes, en particulier : la varicelle, le zona, les feux sauvages, la tuberculose, la scarlatine, l'amygdalite, la pneumonie et les verrues plantaires[60]. De plus, des études ont associés la population possédant les allèles HLA-DRB5\*01 et DRB1\*15:01, qui sont présents dans 15 % de la population générale, à la susceptibilité à la maladie de Parkinson[60, 101]. Outre les changements qualitatifs au sein du répertoire peptidique pouvant être stabilisé sur les molécules du CMH, le niveau d'expression et la stabilité des molécules HLA à la surface des cellules se sont avérés être associés à des maladies[60]. Par exemple, un niveau élevé d'expression de HLA-C augmente le risque de la maladie de Crohn mais favorise également le contrôle de l'infection par le VIH[60]. L'instabilité de HLA-DQ2 et HLA-DQ8 pourrait aussi augmenter le risque de diabète de type 1 alors que la stabilité de HLA-DQ6 pourrait conférer une protection[60].

L'importance de la description, de la quantification et de la compréhension de la complexité de l'immunopectidome et de ses implications dans certaines maladies génétiques est un défi central et fondamental de l'immunologie, avec des implications cliniques importantes dans la médecine de précision. C'est pour cette raison que de nombreux efforts sont en cours pour organiser des bases de données qui caractérisent l'immunopectidome de patients sains ou malades[97]. Dans un avenir rapproché, il n'est pas impossible d'imaginer que toutes les maladies seront caractérisées au niveau de leur effets sur la nature et l'abondance des peptides associés au CMH. De telles connaissances pourraient permettre la mise sur pied de nouvelles stratégies pour le développement d'immunothérapies ou vaccins contre les maladies auto-immunes, infectieuses et contre les cancers[60, 97, 98].

## **5.2 - Le manque de contrôle de la qualité des échantillons étudiés en immunopectidomique**

Les experts de ce domaine s'accordent pour dire que le principal obstacle à l'obtention de données standardisé en l'immunopectidomique est la procédure impliquée dans l'isolation des peptides liés au CMH. Dans d'autres domaines comme la génomique et transcriptomique, les protocoles de préparation des échantillons permettent souvent la possibilité de réaliser toutes les étapes de traitement des échantillons dans un seul tube, ce qui améliore la sensibilité et réduit les pertes au niveau des échantillons. Cependant, en immunopectidomique, les méthodes de purification par affinité et par élution acide douce nécessitent généralement plusieurs étapes, des quantités relativement importantes de matériel biologique et d'anticorps, et, ont des rendements plutôt faibles. Le développement de standards et de méthodes de contrôle de la qualité et de la spécificité dans l'isolation des peptides associés au CMH est crucial afin de permettre une harmonisation des données. De plus, en l'absence d'un protocole unifié pour travailler avec des échantillons d'immunopectidome à travers différents laboratoires, la communauté scientifique est confrontée à l'impossibilité de combiner ou de comparer les données des différentes expérimentations[96].

### **5.2.1 - L'importance de la standardisation des méthodes d'isolation et d'analyses des peptides associés au CMH : collecter, harmoniser et partager les données immunopeptidomiques générées par la communauté**

Pour le futur de l'immunopeptidomique, nous prévoyons que le développement des technologies aboutira à une uniformisation des protocoles. L'harmonisation des méthodes d'isolation des peptides du CMH permettra l'étude de dizaines de milliers d'échantillons et fournira une nouvelle couche d'informations sur la susceptibilité aux maladies, ainsi qu'une meilleure compréhension des associations entre les allèles HLA et les maladies humaines. Ces technologies pourraient fournir des informations essentielles pour améliorer la qualité de vie des patients, pour mieux comprendre les associations entre les allèles HLA et les maladies humaines, ou entre les allèles HLA et la réactivité à certains traitements [60]. Pour se faire, il faudra quantifier l'immunopeptidome de grandes cohortes de population humaine et établir un lien statistique entre les variations au sein de l'immunopeptidome et différents résultats cliniques. De telles études fourniraient des informations très précises sur la prédisposition d'un individu à contracter des maladies immunitaires, ce qui pourrait grandement contribuer aux efforts de prévention sur lesquels le futur de la médecine doit s'appuyer.

Toutefois, pour rendre ces études possibles, plusieurs interrogations devront être résolues. Afin d'évaluer correctement la reproductibilité intra et inter-laboratoire des méthodes de purification par immuno-affinité, des études multi-laboratoire dans lesquelles un ensemble de complexes peptide-CMH pourrait être distribué dans différents groupes de recherche pour la quantification précise du rendement de la procédure. Les résultats d'un tel projet permettraient de clarifier l'incertitude concernant le rendement de la procédure d'isolement des peptides et de déterminer la robustesse de la méthode parmi différents groupes [96].

Il a été démontré que la méthode d'acquisition la plus récente en spectrométrie de masse, connue sous le nom d'acquisition indépendante des données (DIA-MS), est hautement reproductible pour la mesure des peptides associés aux molécules du

complexe majeur d'histocompatibilité, et ce, à travers l'analyse de plusieurs échantillons[54]. Une fois cette méthode développée et déployée par la communauté, elle pourrait représenter une stratégie puissante pour soutenir la mesure cohérente et reproductible de cohortes d'échantillons immunopeptidomiques à grande échelle à travers les utilisateurs et institutions qui en feront l'utilisation. Pour rendre cet objectif une possibilité, la communauté des scientifiques du domaine de l'immunopeptidomique devra établir des lignes directrices et mettre sur pied un rapport sur les informations minimales qui devraient être pris en compte et enregistrés lors de toutes expériences sur le sujet. Un tel rapport décrirait des informations spécifiques sur le type et la préparation des échantillons, l'organisme, le type de tissu et de cellule utilisés, la quantité d'échantillon, les conditions de culture, l'état de la maladie, le typage HLA du modèle utilisé ainsi que l'anticorps utilisé pour l'isolement des peptides.

Au niveau du partage des données, SystemMHC Atlas (<https://systemhcatlas.org>), est une banque de données ouverte en immunopeptidomique dans lequel les résultats d'analyse en spectrométrie de masse sont correctement annotés, stockés et partagés sans restriction. Compte tenu du nombre de bases de données supplémentaires liées à l'immunopeptidomique qui pourraient être développées à l'avenir, il sera primordial de développer des approches pour assurer l'interopérabilité des bases de données afin d'assurer une harmonisation des résultats en SM parmi plusieurs laboratoires[96-98].

La cartographie complète et la quantification des immunopeptidomes devraient devenir aussi sensibles et reproductibles qu'elle l'est actuellement pour d'autres technologies à haut débit révolutionnaires dans les sciences de la vie, telles que le séquençage de l'ADN et de l'ARN de nouvelle génération. La combinaison de procédures d'isolation des peptides robustes et évolutives, de techniques de spectrométrie de masse plus sensibles et reproductibles, ainsi que d'outils logiciels de référence pour l'identification précise des peptides associés au système HLA devrait accroître l'accessibilité et l'impact de l'immunopeptidomique au profit de tous.

### 5.2.2 - Stratégies pour l'optimisation de la reproductibilité dans l'étude de l'immunopeptidomique inter-établissements.

La qualité des données est primordiale afin d'atteindre une répétabilité et reproductibilité à travers plusieurs établissements et instruments d'analyse. À cet égard, il est important d'évaluer et de contrôler la qualité des données immunopeptidomiques obtenues par SM. Dans le domaine de la génomique et de la protéomique, l'importance du contrôle de la qualité et de l'assurance de la qualité est évidente et reconnue depuis longtemps[144]. Différents types d'échantillons ainsi que des solutions logicielles de contrôle de la qualité ont été largement développés et appliqués au fil des ans. Au contraire, les outils de contrôle et d'assurance de la qualité dans le domaine de l'immunopeptidomique basée sur la spectrométrie de masse restent peu documentés en dépit de leur importance pour un développement thérapeutique réussi[58, 117].

Mis à part les travaux de Kovalchik et al (2021) pour la mise au pied du logiciel MHCVizpipe, deux autres études s'étaient penchées sur des mesures du contrôle de la qualité des données générés par spectrométrie de masse. Ghosh et al. (2020)[58] décrivent l'importance de l'exactitude de la masse des peptides obtenus, la concordance des temps de rétention observés et calculés, la répétabilité des temps de rétention et des intensités de signal des peptides détectés. De plus, l'utilisation de peptides synthétiques permet de déterminer la limite de détection des ligands HLA d'intérêt[58]. Dans cette étude, les auteurs se penchent sur certains critères pour valider l'exactitude, la précision, la spécificité, la limite de détection et la robustesse de manière à être conforme aux directives de la FDA et de l'EMA (*Federal Drug Administration et European Medicines Agency*). Pour l'exactitude, l'écart de la masse médiane par rapport à la masse théorique de tous les peptides naturels et synthétiques identifiés doit être inférieure ou égale à 2ppm, pour la précision, Le nombre de peptides identifiés doit satisfaire aux critères d'acceptation de  $\pm 10\%$  de l'écart-type de la répétabilité le premier jour et après une semaine[58]. Au niveau de la spécificité, l'écart-type entre l'ion précurseur et les cinq masses de fragment en tête de ligne des peptides sélectionnés doit être plus petit ou égal à 0,001 Da[58]. Le peptide doit aussi être identifié dans deux des trois répliques. Pour la

limite de détection, ils fixent la limite de détection à la concentration de peptides qui permet d'identifier au moins 50 % des peptides par réplicas, avec un taux de récupération de  $80 \% \pm 20 \%$ [58]. Dans la deuxième étude de CQ, Fritsche et al. (2021)[142] présentent des statistiques permettant de discriminer les vrais ligands HLA des fragments protéolytiques co-isolés indépendants du HLA [142].

Pour permettre la normalisation et l'harmonisation des données entre les différentes plateformes de SM à travers différentes institutions, chaque plateforme d'instrumentation de spectrométrie de masse doit être validée indépendamment en fonction de ses propres spécifications. Pour la validation, les spécifications doivent être adaptées aussi étroitement que possible à la performance optimale des systèmes respectifs[58]. Après la validation d'une plate-forme de SM, une maintenance fréquente du système doit être observée pour surveiller de manière continue la performance de l'instrument à travers différents intervalles fixes. Des peptides de contrôle identifiés de manière fiable peuvent être sélectionnés à cette fin. L'identification et le temps de rétention de ces peptides peut être utilisés pour vérifier de façon routinière les performances de l'analyse par SM. Les peptides de contrôle sélectionnés peuvent être standardisés entre différents systèmes de SM, et ce, à travers plusieurs laboratoires[58].

### **5.3 - L'importance de la réponse T dans le contrôle de l'infection à SRAS-CoV2 par vaccination**

Depuis que l'Organisation mondiale de la santé a déclaré l'écllosion de Covid-19 comme étant pandémique le 11 mars 2020, la communauté scientifique mondiale a répondu à cette urgence de plusieurs manières ingénieuses en réorientant des approches et des plates-formes vaccinales non approuvées jusqu'alors et initialement destinées à des applications en immunologie, ayant comme résultat le développement de vaccins contre le COVID-19 à un rythme sans précédent. Cependant, il n'existe toujours pas de vaccins appropriés permettant d'induire spécifiquement une réponse des cellules T. Actuellement, pour la plupart des vaccins approuvés, les évaluations de la protection se sont principalement concentrées sur les réponses immunitaires humorales[167], ne

fournissant que des données très limitées sur l'induction des lymphocytes T. Toutefois, ces vaccins peuvent se révéler moins efficace pour les tranches de population car ils pourraient ne montrer qu'une efficacité limitée ou nécessiter une induction supplémentaire des lymphocytes T chez les patients âgés, cancéreux ou traités par des régimes de médicaments qui réduisent la production d'anticorps[95]. Les trois composants fondamentaux du système immunitaire adaptatif : les cellules B, les cellules T CD4+ et les cellules T CD8+, jouent des rôles différents dans différentes infections virales et dans les vaccins. Il est donc essentiel d'étudier directement l'immunité adaptative au SRAS-CoV-2 pour comprendre comment parer aux mécanismes d'infection de COVID-19. Des preuves ont commencé à révéler que les cellules CD4+T, les cellules CD8+T et les anticorps neutralisants contribuent tous au contrôle du SRAS-CoV-2 dans les cas non hospitalisés et hospitalisés de COVID-19[95, 168]

La compréhension des mécanismes d'évasion immunitaire d'un virus est souvent importante pour saisir la pathogenèse de celui-ci. Dans le cas du SRAS-CoV-2, le virus est exceptionnellement efficace pour échapper au déclenchement des réponses immunitaires innées précoces[169]. Il est possible qu'une grande partie de la virulence et de la gravité des manifestations clinique du COVID-19 soit une conséquence de ce tour de force du SRAS-CoV-2 dans l'évasion de l'immunité innée[168, 169]. Conjointement aux fonctions du système inné qui crée un état antiviral dans l'environnement tissulaire, une des fonctions du système immunitaire inné consiste à amorcer la réponse immunitaire adaptative. La réponse immunitaire adaptative met généralement du temps à générer assez de cellules pour contrôler une infection virale, environ 6 à 10 jours après l'amorçage par le système inné[169]. Au cours de l'infection par le SRAS-CoV-2, *in-vitro* et chez l'humain, le virus est particulièrement efficace pour éviter ou retarder le déclenchement des réponses immunitaires innées associées à la production d'interféron[169]. En l'absence de ces réponses, le virus se reproduit initialement sans relâche et les réponses immunitaires adaptatives ne sont pas amorcées tant que les cellules du systèmes immunitaires innées ne sont pas activées[169]. Les multiples facettes des manifestations cliniques du COVID-19 dépendent donc en partie de l'harmonisation des fonctions et

activations des cellules effectrices du système adaptatif [168, 169]. En effet, des études menées sur des patients atteints de COVID-19 en phase aiguë et en convalescence ont montré que les réponses des lymphocytes T spécifiques du SRAS-CoV-2 sont significativement associées à une maladie moins grave, suggérant que les réponses des lymphocytes T peuvent être importantes pour le contrôle et la résolution d'une infection primaire par le SRAS-CoV-2.[168, 170]. Il est possible que, dans les cas où la réponse T est insuffisante, le système immunitaire inné tente de combler les lacunes laissés par l'absence de réponse des lymphocytes T en essayant de contrôler le virus par une réponse immunitaire innée plus étendue. Cette solution s'avère intenable, car une réponse innée massive entraîne une immunopathologie pulmonaire excessive et cause de nombreuses lésions au niveau des poumons[171].

Un vaccin composé de peptides dérivés du SRAS-CoV-2 identifiés par les méthodes décrites dans cet ouvrage serait particulièrement approprié pour les personnes qui ont tendance à avoir une activité plus faible des lymphocytes T CD4, par exemple, les patients atteints de tumeurs malignes à cellules B[95]. Malgré qu'un tel vaccin constitué uniquement d'épitopes de lymphocytes T ne prévienne pas complètement l'infection comme c'est le cas pour les vaccins traditionnels visant à obtenir une forte réponse en anticorps neutralisants, il aiderait le système immunitaire du patient à résoudre rapidement l'infection naissante en favorisant une production plus rapide d'anticorps, rendue possible par les cellules T CD4+ induites par le vaccin et un meilleur contrôle de l'infection subséquente par une production adéquate de cellules T CD8+. L'induction des lymphocytes T contre plusieurs protéines virales devrait également présenter l'avantage d'une protection contre les mutations virales n'affectant qu'une seule protéine virale. Dans le cas du SRAS-CoV-2, ce type de vaccin offrirait théoriquement une protection contre les nouveaux sérotypes et variants viraux inconnus ou qui pourraient émerger à court, moyen et long terme[95, 168].

#### **5.4 - La prochaine génération en immunopectidomique : prédire, traiter, faire le suivi et guérir les cancers, maladies auto-immunes et infectieuses.**

En suivant la voie des études du génome, du transcriptome, du protéome et du métabolome, les technologies capables d'effectuer des mesures robustes et complètes des immunopectidomes à l'échelle d'une population seraient extrêmement puissantes car elles fourniraient des preuves physiques directes de l'identité et de la quantité des peptides qui interagissent et engagent l'action des cellules T. Ces technologies pourraient fournir des informations essentielles pour mieux comprendre les associations entre les allèles HLA et les maladies humaines ou entre les allèles HLA et la réponse au traitement. Dans ce contexte, nous prévoyons que le développement des technologies immunopectidomiques conduiront à terme des études d'associations à l'échelle de l'immunopectidome et fourniront une nouvelle couche d'informations sur la susceptibilité de certains loci envers différentes maladies liées ou non au système immunitaire ainsi que l'efficacité de la réponse à des vaccins ou immunothérapies.

À cet égard, l'immunopectidomique deviendra un volet central de l'immunologie pour prédire, diagnostiquer, surveiller et traiter les maladies immunitaires et aura donc un impact profond sur la médecine moderne[60].

## Conclusion

La spectrométrie de masse est une approche puissante pour l'analyse des peptides qui composent l'immunopeptidome d'un organisme. La sensibilité et fiabilité de cet outil fait actuellement de cette méthode d'analyse une des approches les plus précises et prometteuse dans l'identification et la caractérisation de peptides aberrants ou de provenance exogène qui interagissent avec les molécules du complexe majeur d'histocompatibilité.

Toutefois, comme démontré dans cette étude, ce domaine souffre de lacunes au niveau de l'harmonisation des protocoles, des données générées, ainsi qu'un manque d'outils de contrôle et d'assurance de la qualité des échantillons.

Dans cette expérimentation nous avons évalué une approche de contrôle de qualité pour la caractérisation de l'immunopeptidome de lignées cellulaires infectées par un virus de la famille des coronavirus. Nous avons démontré que le protocole d'isolation des peptides utilisé est efficace et reproductible entre les expérimentations et utilisateurs d'une même institution. Toutefois, nous avons conclu que l'utilisation de différents protocoles d'isolation des peptides ainsi que logiciels d'interprétation des données brutes en spectrométrie de masse peuvent mener à des divergences majeures au niveau de la reproductibilité des résultats.

L'utilisation de MVP, un outil logiciel de contrôle de qualité semi-automatisé, permet l'évaluation rapide et simultanée de multiples ensembles de données immunopeptidomiques et fournit une évaluation rapide de la qualité, de la composition et de la spécificité du CMH des échantillons analysés, ce qui accélère considérablement le processus décisionnel de contrôle de qualité en vue de l'interprétation des données.

Le modèle d'étude composé des cellules MRC5 et du Coronavirus OC-43 a permis l'identification de 9 peptides viraux à la surface des molécules du CMH-1 des cellules analysées malgré l'utilisation d'un échantillon considérablement réduit. Ce travail démontre qu'un utilisateur avec peu d'expérience est en mesure, avec des ressources limitées, de faire la culture des cellules, de procéder à l'infection, d'isoler les peptides du

CMH, de les analysés par spectrométrie de masse et de procéder à l'interprétation des résultats à l'aide d'outils logiciels de contrôle de la qualité.

Les méthodes d'analyse utilisées dans cet ouvrage peuvent aider à l'harmonisation et au partage des données immuno-peptidomiques générées par la communauté.

Également, les résultats de ce travail peuvent aider à identifier des peptides pertinents dans la conception de vaccins de seconde génération contre les coronavirus basés sur la réponse T. De tels vaccins élicitant une réponse contre plusieurs portions du génome viral pourraient être potentiellement plus efficace au niveau de la longévité en abaissant les chances du virus de contourner sa protection tout en étant moins dangereux pour certains patients à risque comparativement aux vaccins actuels qui élicitent une forte réponse humorale.

En utilisant l'approche utilisé dans cet ouvrage sur plusieurs modèles cellulaires infectés par différents coronavirus et en identifiant des signatures génétiques retrouvées chez plusieurs membres de la grande famille des coronavirus à la surface des molécules du CMH, il serait possible de concevoir des vaccins élicitant une réponse contre de nombreux coronavirus et ainsi potentiellement réduire les chances de subir les conséquences d'une autre pandémie mondiale reliée à l'émergence d'un coronavirus.

## Bibliographie

1. Litman, G.W., J.P. Rast, and S.D. Fugmann, *The origins of vertebrate adaptive immunity*. Nat Rev Immunol, 2010. **10**(8): p. 543-53.
2. Mariuzza, R.A., et al., *Structural insights into the evolution of the adaptive immune system: the variable lymphocyte receptors of jawless vertebrates*. Biol Chem, 2010. **391**(7): p. 753-60.
3. Parkin, J. and B. Cohen, *An overview of the immune system*. Lancet, 2001. **357**(9270): p. 1777-89.
4. Kumar, H., T. Kawai, and S. Akira, *Pathogen recognition by the innate immune system*. Int Rev Immunol, 2011. **30**(1): p. 16-34.
5. Pancer, Z. and M.D. Cooper, *The evolution of adaptive immunity*. Annu Rev Immunol, 2006. **24**: p. 497-518.
6. Orkin, S.H., *Diversification of haematopoietic stem cells to specific lineages*. Nat Rev Genet, 2000. **1**(1): p. 57-64.
7. Wilson, A. and A. Trumpp, *Bone-marrow haematopoietic-stem-cell niches*. Nat Rev Immunol, 2006. **6**(2): p. 93-106.
8. Smith-Garvin, J.E., G.A. Koretzky, and M.S. Jordan, *T cell activation*. Annu Rev Immunol, 2009. **27**: p. 591-619.
9. Kurosaki, T., H. Shinohara, and Y. Baba, *B cell signaling and fate decision*. Annu Rev Immunol, 2010. **28**: p. 21-55.
10. Blank, U. and S. Karlsson, *TGF-beta signaling in the control of hematopoietic stem cells*. Blood, 2015. **125**(23): p. 3542-50.
11. Schatz, D.G. and Y. Ji, *Recombination centres and the orchestration of V(D)J recombination*. Nat Rev Immunol, 2011. **11**(4): p. 251-63.
12. Starr, T.K., S.C. Jameson, and K.A. Hogquist, *Positive and negative selection of T cells*. Annu Rev Immunol, 2003. **21**: p. 139-76.
13. Miller, J., *The function of the thymus and its impact on modern medicine*. Science, 2020. **369**(6503).
14. Chen, L. and D.B. Flies, *Molecular mechanisms of T cell co-stimulation and co-inhibition*. Nat Rev Immunol, 2013. **13**(4): p. 227-42.
15. Figueiredo, M., R. Zilhao, and H. Neves, *Thymus Inception: Molecular Network in the Early Stages of Thymus Organogenesis*. Int J Mol Sci, 2020. **21**(16).
16. Ohshima, Y. and G. Delespesse, *T cell-derived IL-4 and dendritic cell-derived IL-12 regulate the lymphokine-producing phenotype of alloantigen-primed naive human CD4 T cells*. J Immunol, 1997. **158**(2): p. 629-36.
17. Vacafloras, A., et al., *Exposure of Human CD4 T Cells to IL-12 Results in Enhanced TCR-Induced Cytokine Production, Altered TCR Signaling, and Increased Oxidative Metabolism*. PLoS One, 2016. **11**(6): p. e0157175.
18. Forbes, E., et al., *Differential requirements for IL-4/STAT6 signalling in CD4 T-cell fate determination and Th2-immune effector responses*. Immunol Cell Biol, 2010. **88**(3): p. 240-3.
19. Ruterbusch, M., et al., *In Vivo CD4(+) T Cell Differentiation and Function: Revisiting the Th1/Th2 Paradigm*. Annu Rev Immunol, 2020. **38**: p. 705-725.
20. de Verteuil, D., et al., *Origin and plasticity of MHC I-associated self peptides*. Autoimmun Rev, 2012. **11**(9): p. 627-35.

21. van den Elsen, P.J., et al., *Transcriptional regulation of antigen presentation*. *Curr Opin Immunol*, 2004. **16**(1): p. 67-75.
22. Corse, E., R.A. Gottschalk, and J.P. Allison, *Strength of TCR-peptide/MHC interactions and in vivo T cell responses*. *J Immunol*, 2011. **186**(9): p. 5039-45.
23. den Haan, J.M., R. Arens, and M.C. van Zelm, *The activation of the adaptive immune system: cross-talk between antigen-presenting cells, T cells and B cells*. *Immunol Lett*, 2014. **162**(2 Pt B): p. 103-12.
24. Kolumam, G.A., et al., *Type I interferons act directly on CD8 T cells to allow clonal expansion and memory formation in response to viral infection*. *J Exp Med*, 2005. **202**(5): p. 637-50.
25. Sprent, J. and C.D. Surh, *T cell memory*. *Annu Rev Immunol*, 2002. **20**: p. 551-79.
26. Gebhardt, T. and L.K. Mackay, *Local immunity by tissue-resident CD8(+) memory T cells*. *Front Immunol*, 2012. **3**: p. 340.
27. Hofmann, M., et al., *Thymus-resident memory CD8+ T cells mediate local immunity*. *Eur J Immunol*, 2013. **43**(9): p. 2295-304.
28. Jakubzick, C.V., G.J. Randolph, and P.M. Henson, *Monocyte differentiation and antigen-presenting functions*. *Nat Rev Immunol*, 2017. **17**(6): p. 349-362.
29. Kratky, W., et al., *Direct activation of antigen-presenting cells is required for CD8+ T-cell priming and tumor vaccination*. *Proc Natl Acad Sci U S A*, 2011. **108**(42): p. 17414-9.
30. Reeves, E. and E. James, *Antigen processing and immune regulation in the response to tumours*. *Immunology*, 2017. **150**(1): p. 16-24.
31. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II antigen presentation*. *Nat Rev Immunol*, 2011. **11**(12): p. 823-36.
32. Bjorkman, P.J., et al., *Structure of the human class I histocompatibility antigen, HLA-A2*. *Nature*, 1987. **329**(6139): p. 506-12.
33. Blum, J.S., P.A. Wearsch, and P. Cresswell, *Pathways of antigen processing*. *Annu Rev Immunol*, 2013. **31**: p. 443-73.
34. Shastri, N., et al., *All the peptides that fit: the beginning, the middle, and the end of the MHC class I antigen-processing pathway*. *Immunol Rev*, 2005. **207**: p. 31-41.
35. Parham, P., E.J. Adams, and K.L. Arnett, *The origins of HLA-A,B,C polymorphism*. *Immunol Rev*, 1995. **143**: p. 141-80.
36. Rammensee, H.G., *Chemistry of peptides associated with MHC class I and class II molecules*. *Curr Opin Immunol*, 1995. **7**(1): p. 85-96.
37. Falk, K., et al., *Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules*. *Nature*, 1991. **351**(6324): p. 290-6.
38. DiBrino, M., et al., *HLA-A1 and HLA-A3 T cell epitopes derived from influenza virus proteins predicted from peptide binding motifs*. *J Immunol*, 1993. **151**(11): p. 5930-5.
39. Sidney, J., et al., *HLA class I supertypes: a revised and updated classification*. *BMC Immunol*, 2008. **9**: p. 1.
40. Robinson, J., et al., *Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles*. *PLoS Genet*, 2017. **13**(6): p. e1006862.

41. Shraibman, B., et al., *Human Leukocyte Antigen (HLA) Peptides Derived from Tumor Antigens Induced by Inhibition of DNA Methylation for Development of Drug-facilitated Immunotherapy*. Mol Cell Proteomics, 2016. **15**(9): p. 3058-70.
42. Silver, M.L., et al., *Atomic structure of a human MHC molecule presenting an influenza virus peptide*. Nature, 1992. **360**(6402): p. 367-9.
43. Madden, D.R., *The three-dimensional structure of peptide-MHC complexes*. Annu Rev Immunol, 1995. **13**: p. 587-622.
44. Hunt, D.F., et al., *Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry*. Science, 1992. **255**(5049): p. 1261-3.
45. Haga, S.W. and H.F. Wu, *Overview of software options for processing, analysis and interpretation of mass spectrometric proteomic data*. J Mass Spectrom, 2014. **49**(10): p. 959-69.
46. Andreatta, M., B. Alvarez, and M. Nielsen, *GibbsCluster: unsupervised clustering and alignment of peptide sequences*. Nucleic Acids Res, 2017. **45**(W1): p. W458-W463.
47. Jurtz, V., et al., *NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data*. J Immunol, 2017. **199**(9): p. 3360-3368.
48. Alvarez, B., et al., *Computational Tools for the Identification and Interpretation of Sequence Motifs in Immunopeptidomes*. Proteomics, 2018. **18**(12): p. e1700252.
49. Andreatta, M., et al., *MS-Rescue: A Computational Pipeline to Increase the Quality and Yield of Immunopeptidomics Experiments*. Proteomics, 2019. **19**(4): p. e1800357.
50. Mei, S., et al., *A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction*. Brief Bioinform, 2020. **21**(4): p. 1119-1135.
51. Reynisson, B., et al., *NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data*. Nucleic Acids Res, 2020. **48**(W1): p. W449-W454.
52. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*. Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.
53. Domon, B. and R. Aebersold, *Challenges and opportunities in proteomics data analysis*. Mol Cell Proteomics, 2006. **5**(10): p. 1921-6.
54. Caron, E., et al., *Analysis of Major Histocompatibility Complex (MHC) Immunopeptidomes Using Mass Spectrometry*. Mol Cell Proteomics, 2015. **14**(12): p. 3105-17.
55. Devabhaktuni, A. and J.E. Elias, *Application of de Novo Sequencing to Large-Scale Complex Proteomics Data Sets*. J Proteome Res, 2016. **15**(3): p. 732-42.
56. Bassani-Sternberg, M., *Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens*. Methods Mol Biol, 2018. **1719**: p. 209-221.
57. Purcell, A.W., S.H. Ramarathinam, and N. Ternette, *Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics*. Nat Protoc, 2019. **14**(6): p. 1687-1707.

58. Ghosh, M., et al., *Guidance Document: Validation of a High-Performance Liquid Chromatography-Tandem Mass Spectrometry Immunopeptidomics Assay for the Identification of HLA Class I Ligands Suitable for Pharmaceutical Therapies*. Mol Cell Proteomics, 2020. **19**(3): p. 432-443.
59. Poulos, R.C., et al., *Strategies to enable large-scale proteomics for reproducible research*. Nat Commun, 2020. **11**(1): p. 3793.
60. Vizcaino, J.A., et al., *The Human Immunopeptidome Project: A Roadmap to Predict and Treat Immune Diseases*. Mol Cell Proteomics, 2020. **19**(1): p. 31-49.
61. Ma, Y., et al., *Large-scale identification of T cell epitopes derived from SARS-CoV-2 for the development of peptide vaccines against COVID-19*. J Infect Dis, 2021.
62. Weingarten-Gabbay, S., et al., *Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs*. Cell, 2021. **184**(15): p. 3962-3980 e17.
63. Wieczorek, M., et al., *Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation*. Front Immunol, 2017. **8**: p. 292.
64. Yewdell, J.W., E. Reits, and J. Neefjes, *Making sense of mass destruction: quantitating MHC class I antigen presentation*. Nat Rev Immunol, 2003. **3**(12): p. 952-61.
65. Kleiger, G. and T. Mayor, *Perilous journey: a tour of the ubiquitin-proteasome system*. Trends Cell Biol, 2014. **24**(6): p. 352-9.
66. Bochtler, M., et al., *The proteasome*. Annu Rev Biophys Biomol Struct, 1999. **28**: p. 295-317.
67. de Verteuil, D., et al., *Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules*. Mol Cell Proteomics, 2010. **9**(9): p. 2034-47.
68. Vigneron, N. and B.J. Van den Eynde, *Proteasome subtypes and regulators in the processing of antigenic peptides presented by class I molecules of the major histocompatibility complex*. Biomolecules, 2014. **4**(4): p. 994-1025.
69. Akiyama, K., et al., *cDNA cloning and interferon gamma down-regulation of proteasomal subunits X and Y*. Science, 1994. **265**(5176): p. 1231-4.
70. Heink, S., et al., *IFN-gamma-induced immune adaptation of the proteasome system is an accelerated and transient response*. Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9241-6.
71. Pushpakumara, P.D., et al., *Identification of Novel Candidate CD8(+) T Cell Epitopes of the SARS-CoV2 with Homology to Other Seasonal Coronaviruses*. Viruses, 2021. **13**(6).
72. Neefjes, J.J., F. Momburg, and G.J. Hammerling, *Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter*. Science, 1993. **261**(5122): p. 769-71.
73. Kienast, A., et al., *Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin*. Nat Immunol, 2007. **8**(8): p. 864-72.
74. Jensen, P.E., *Recent advances in antigen processing and presentation*. Nat Immunol, 2007. **8**(10): p. 1041-8.

75. Van Hateren, A., et al., *The cell biology of major histocompatibility complex class I assembly: towards a molecular understanding*. Tissue Antigens, 2010. **76**(4): p. 259-75.
76. Trowsdale, J. and J.C. Knight, *Major histocompatibility complex genomics and human disease*. Annu Rev Genomics Hum Genet, 2013. **14**: p. 301-23.
77. Heath, W.R. and F.R. Carbone, *Cross-presentation, dendritic cells, tolerance and immunity*. Annu Rev Immunol, 2001. **19**: p. 47-64.
78. Heath, W.R. and F.R. Carbone, *Cross-presentation in viral immunity and self-tolerance*. Nat Rev Immunol, 2001. **1**(2): p. 126-34.
79. Adiko, A.C., et al., *Intracellular Transport Routes for MHC I and Their Relevance for Antigen Cross-Presentation*. Front Immunol, 2015. **6**: p. 335.
80. Colbert, J.D., F.M. Cruz, and K.L. Rock, *Cross-presentation of exogenous antigens on MHC I molecules*. Curr Opin Immunol, 2020. **64**: p. 1-8.
81. Vyas, J.M., A.G. Van der Veen, and H.L. Ploegh, *The known unknowns of antigen processing and presentation*. Nat Rev Immunol, 2008. **8**(8): p. 607-18.
82. Joffre, O.P., et al., *Cross-presentation by dendritic cells*. Nat Rev Immunol, 2012. **12**(8): p. 557-69.
83. Rock, K.L., E. Reits, and J. Neefjes, *Present Yourself! By MHC Class I and MHC Class II Molecules*. Trends Immunol, 2016. **37**(11): p. 724-737.
84. Radley, E., et al., *Genomic organization of HLA-DMA and HLA-DMB. Comparison of the gene organization of all six class II families in the human major histocompatibility complex*. J Biol Chem, 1994. **269**(29): p. 18834-8.
85. Jones, E.Y., *MHC class I and class II structures*. Curr Opin Immunol, 1997. **9**(1): p. 75-9.
86. Painter, C.A. and L.J. Stern, *Conformational variation in structures of classical and non-classical MHCII proteins and functional implications*. Immunol Rev, 2012. **250**(1): p. 144-57.
87. Shraibman, B., et al., *Identification of Tumor Antigens Among the HLA Peptidomes of Glioblastoma Tumors and Plasma*. Mol Cell Proteomics, 2019. **18**(6): p. 1255-1268.
88. Wolchok, J., *Putting the Immunologic Brakes on Cancer*. Cell, 2018. **175**(6): p. 1452-1454.
89. Altmann, D.M., *A Nobel Prize-worthy pursuit: cancer immunology and harnessing immunity to tumour neoantigens*. Immunology, 2018. **155**(3): p. 283-284.
90. Creech, A.L., et al., *The Role of Mass Spectrometry and Proteogenomics in the Advancement of HLA Epitope Prediction*. Proteomics, 2018. **18**(12): p. e1700259.
91. Poran, A., et al., *Sequence-based prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes*. Genome Med, 2020. **12**(1): p. 70.
92. Saunders, K.O., et al., *Neutralizing antibody vaccine for pandemic and pre-emergent coronaviruses*. Nature, 2021. **594**(7864): p. 553-559.
93. Joyce, S. and N. Ternette, *Know thy immune self and non-self: Proteomics informs on the expanse of self and non-self, and how and where they arise*. Proteomics, 2021: p. e2000143.

94. Gfeller, D. and M. Bassani-Sternberg, *Predicting Antigen Presentation-What Could We Learn From a Million Peptides?* Front Immunol, 2018. **9**: p. 1716.
95. Rammensee, H.G., et al., *Designing a SARS-CoV-2 T-Cell-Inducing Vaccine for High-Risk Patient Groups*. Vaccines (Basel), 2021. **9**(5).
96. Caron, E., et al., *A Case for a Human Immuno-Peptidome Project Consortium*. Immunity, 2017. **47**(2): p. 203-208.
97. Shao, W., et al., *The SystemMHC Atlas project*. Nucleic Acids Res, 2018. **46**(D1): p. D1237-D1247.
98. Shao, W., et al., *The SystemMHC Atlas: a Computational Pipeline, a Website, and a Data Repository for Immunopeptidomic Analyses*. Methods Mol Biol, 2020. **2120**: p. 173-181.
99. Battle, R. and B. Clark, *Quantitative analysis of human leucocyte antigen expression during culture of Epstein-Barr virus-transformed cell lines using the dako QIFIKIT*. Br J Biomed Sci, 2007. **64**(1): p. 32-4.
100. Urlaub, D. and C. Watzl, *Coated Latex Beads as Artificial Cells for Quantitative Investigations of Receptor/Ligand Interactions*. Curr Protoc Immunol, 2020. **131**(1): p. e111.
101. Kuznetsov, A., et al., *Critical Review of Existing MHC I Immunopeptidome Isolation Methods*. Molecules, 2020. **25**(22).
102. Storkus, W.J., et al., *Identification of T-cell epitopes: rapid isolation of class I-presented peptides from viable cells by mild acid elution*. J Immunother Emphasis Tumor Immunol, 1993. **14**(2): p. 94-103.
103. Fortier, M.H., et al., *The MHC class I peptide repertoire is molded by the transcriptome*. J Exp Med, 2008. **205**(3): p. 595-610.
104. Sturm, T., et al., *Mild Acid Elution and MHC Immunoaffinity Chromatography Reveal Similar Albeit Not Identical Profiles of the HLA Class I Immunopeptidome*. J Proteome Res, 2021. **20**(1): p. 289-304.
105. Hassan, C., et al., *Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes*. J Proteomics, 2014. **109**: p. 240-4.
106. Cole, D.K., *The ultimate mix and match: making sense of HLA alleles and peptide repertoires*. Immunol Cell Biol, 2015. **93**(6): p. 515-6.
107. Mester, G., V. Hoffmann, and S. Stevanovic, *Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands*. Cell Mol Life Sci, 2011. **68**(9): p. 1521-32.
108. Mommen, G.P., et al., *Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD)*. Proc Natl Acad Sci U S A, 2014. **111**(12): p. 4507-12.
109. Tabb, D.L., et al., *Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry*. J Proteome Res, 2010. **9**(2): p. 761-76.
110. Caron, E., et al., *An open-source computational and data resource to analyze digital maps of immunopeptidomes*. Elife, 2015. **4**.
111. Picotti, P. and R. Aebersold, *Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions*. Nat Methods, 2012. **9**(6): p. 555-66.

112. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics*. Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
113. Sajic, T., Y. Liu, and R. Aebersold, *Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications*. Proteomics Clin Appl, 2015. **9**(3-4): p. 307-21.
114. Hickman, H.D., et al., *Toward a definition of self: proteomic evaluation of the class I peptide repertoire*. J Immunol, 2004. **172**(5): p. 2944-52.
115. Patterson, S.D., *Data analysis--the Achilles heel of proteomics*. Nat Biotechnol, 2003. **21**(3): p. 221-2.
116. Bichmann, L., et al., *MHCquant: Automated and Reproducible Data Analysis for Immunopeptidomics*. J Proteome Res, 2019. **18**(11): p. 3876-3884.
117. Kovalchik, K.A., et al., *MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- to Large-Scale Immunopeptidome Data Sets*. Mol Cell Proteomics, 2021: p. 100178.
118. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-67.
119. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database*. J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.
120. Zhang, J., et al., *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification*. Mol Cell Proteomics, 2012. **11**(4): p. M111 010587.
121. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **579**(7798): p. 270-273.
122. Menachery, V.D., R.L. Graham, and R.S. Baric, *Jumping species-a mechanism for coronavirus persistence and survival*. Curr Opin Virol, 2017. **23**: p. 1-7.
123. Wec, A.Z., et al., *Broad neutralization of SARS-related viruses by human monoclonal antibodies*. Science, 2020. **369**(6504): p. 731-736.
124. Grifoni, A., et al., *SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19*. Cell Host Microbe, 2021. **29**(7): p. 1076-1092.
125. Li, D., et al., *The functions of SARS-CoV-2 neutralizing and infection-enhancing antibodies in vitro and in mice and nonhuman primates*. bioRxiv, 2021.
126. Dykema, A.G., et al., *Functional characterization of CD4+ T cell receptors crossreactive for SARS-CoV-2 and endemic coronaviruses*. J Clin Invest, 2021. **131**(10).
127. Martinez, D.R., et al., *A broadly cross-reactive antibody neutralizes and protects against sarbecovirus challenge in mice*. Sci Transl Med, 2021: p. eabj7125.
128. Geers, D., et al., *SARS-CoV-2 variants of concern partially escape humoral but not T-cell responses in COVID-19 convalescent donors and vaccinees*. Sci Immunol, 2021. **6**(59).
129. Zohar, T., et al., *Compromised Humoral Functional Evolution Tracks with SARS-CoV-2 Mortality*. Cell, 2020. **183**(6): p. 1508-1519 e12.

130. Tarke, A., et al., *Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases*. Cell Rep Med, 2021. **2**(2): p. 100204.
131. Liu, G., et al., *Computationally Optimized SARS-CoV-2 MHC Class I and II Vaccine Formulations Predicted to Target Human Haplotype Distributions*. Cell Syst, 2020. **11**(2): p. 131-144 e6.
132. Sarkizova, S., et al., *A large peptidome dataset improves HLA class I epitope prediction across most of the human population*. Nat Biotechnol, 2020. **38**(2): p. 199-209.
133. Zedan, H.T. and G.K. Nasrallah, *Is preexisting immunity to seasonal coronaviruses limited to cross-reactivity with SARS-CoV-2? A seroprevalence cross-sectional study in north-eastern France*. EBioMedicine, 2021. **71**: p. 103580.
134. Kim, D.E., et al., *Natural Bis-Benzylisoquinoline Alkaloids-Tetrandrine, Fangchinoline, and Cepharanthine, Inhibit Human Coronavirus OC43 Infection of MRC-5 Human Lung Cells*. Biomolecules, 2019. **9**(11).
135. Sirois, I., et al., *Immunopeptidomics: Isolation of Mouse and Human MHC Class I- and II-Associated Peptides for Mass Spectrometry Analysis*. J Vis Exp, 2021(176).
136. Hassan, C., et al., *The human leukocyte antigen-presented ligandome of B lymphocytes*. Mol Cell Proteomics, 2013. **12**(7): p. 1829-43.
137. Hulspas, R., *Titration of fluorochrome-conjugated antibodies for labeling cell surface markers on live cells*. Curr Protoc Cytom, 2010. **Chapter 6**: p. Unit 6 29.
138. Bentley, G., et al., *High-resolution, high-throughput HLA genotyping by next-generation sequencing*. Tissue Antigens, 2009. **74**(5): p. 393-403.
139. Berlin, C., et al., *Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy*. Leukemia, 2016. **30**(4): p. 1003-4.
140. Feng, F., et al., *Establishment and characterization of 38 novel patient-derived primary cancer cell lines using multi-region sampling revealing intra-tumor heterogeneity of gallbladder carcinoma*. Hum Cell, 2021. **34**(3): p. 918-931.
141. Kobayashi, K.S. and P.J. van den Elsen, *NLRC5: a key regulator of MHC class I-dependent immune responses*. Nat Rev Immunol, 2012. **12**(12): p. 813-20.
142. Fritsche, J., et al., *Pitfalls in HLA Ligandomics-How to Catch a Li(e)gand*. Mol Cell Proteomics, 2021. **20**: p. 100110.
143. Zhang, L., et al., *Automated Ligand Purification Platform Accelerates Immunopeptidome Analysis by Mass Spectrometry*. J Proteome Res, 2021. **20**(1): p. 393-408.
144. Bittremieux, W., et al., *Quality control in mass spectrometry-based proteomics*. Mass Spectrom Rev, 2018. **37**(5): p. 697-711.
145. Guo, Y., et al., *Three-stage quality control strategies for DNA re-sequencing data*. Brief Bioinform, 2014. **15**(6): p. 879-89.
146. Guo, Y., et al., *Multi-perspective quality control of Illumina exome sequencing data using QC3*. Genomics, 2014. **103**(5-6): p. 323-8.
147. O'Donnell, T.J., et al., *MHCflurry: Open-Source Class I MHC Binding Affinity Prediction*. Cell Syst, 2018. **7**(1): p. 129-132 e4.

148. O'Donnell, T. and A. Rubinsteyn, *High-Throughput MHC I Ligand Prediction Using MHCflurry*. *Methods Mol Biol*, 2020. **2120**: p. 113-127.
149. Racle, J., et al., *Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes*. *Nat Biotechnol*, 2019. **37**(11): p. 1283-1286.
150. Marcu, A., et al., *HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy*. *J Immunother Cancer*, 2021. **9**(4).
151. Lex, A., et al., *UpSet: Visualization of Intersecting Sets*. *IEEE Trans Vis Comput Graph*, 2014. **20**(12): p. 1983-92.
152. Blais, M.E., T. Dong, and S. Rowland-Jones, *HLA-C as a mediator of natural killer and T-cell activation: spectator or key player?* *Immunology*, 2011. **133**(1): p. 1-7.
153. Mulligan, M.J., et al., *Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults*. *Nature*, 2020. **586**(7830): p. 589-593.
154. Ledford, H., *How 'killer' T cells could boost COVID immunity in face of new variants*. *Nature*, 2021. **590**(7846): p. 374-375.
155. Callaway, E. and S. Mallapaty, *Novavax offers first evidence that COVID vaccines protect people against variants*. *Nature*, 2021. **590**(7844): p. 17.
156. Jean, A., et al., *Severity and outcome associated with human coronavirus OC43 infections among children*. *Pediatr Infect Dis J*, 2013. **32**(4): p. 325-9.
157. Nelde, A., et al., *SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition*. *Nat Immunol*, 2021. **22**(1): p. 74-85.
158. Li, M., et al., *Rational Design of a Pan-Coronavirus Vaccine Based on Conserved CTL Epitopes*. *Viruses*, 2021. **13**(2).
159. Guo, H.C., et al., *Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle*. *Nature*, 1992. **360**(6402): p. 364-6.
160. Hirose, R., et al., *A Cytopathic Effect-Based Tissue Culture Method for HCoV-OC43 Titration Using TMPRSS2-Expressing VeroE6 Cells*. *mSphere*, 2021. **6**(3).
161. Feero, W.G., *Introducing "Genomics and Precision Health"*. *JAMA*, 2017. **317**(18): p. 1842-1843.
162. Dendrou, C.A., et al., *HLA variation and disease*. *Nat Rev Immunol*, 2018. **18**(5): p. 325-339.
163. Gough, S.C. and M.J. Simmonds, *The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action*. *Curr Genomics*, 2007. **8**(7): p. 453-65.
164. Faridi, P., R. Aebbersold, and E. Caron, *A first dataset toward a standardized community-driven global mapping of the human immunopeptidome*. *Data Brief*, 2016. **7**: p. 201-5.
165. Gonzalez-Angulo, A.M., B.T. Hennessy, and G.B. Mills, *Future of personalized medicine in oncology: a systems biology approach*. *J Clin Oncol*, 2010. **28**(16): p. 2777-83.
166. Trowsdale, J., *The MHC, disease and selection*. *Immunol Lett*, 2011. **137**(1-2): p. 1-8.
167. Kellam, P. and W. Barclay, *The dynamics of humoral immune responses following SARS-CoV-2 infection and the potential for reinfection*. *J Gen Virol*, 2020. **101**(8): p. 791-797.

168. Sette, A. and S. Crotty, *Adaptive immunity to SARS-CoV-2 and COVID-19*. Cell, 2021. **184**(4): p. 861-880.
169. Blanco-Melo, D., et al., *Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19*. Cell, 2020. **181**(5): p. 1036-1045 e9.
170. Liao, M., et al., *Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19*. Nat Med, 2020. **26**(6): p. 842-844.
171. Li, S., et al., *Clinical and pathological investigation of patients with severe COVID-19*. JCI Insight, 2020. **5**(12).

# **Annexe 1 – Matériel et horaire optimal pour culture cellulaire à grande échelle.**

Cet horaire détaillé a été mis sur pieds, pour récolter de manière optimale, des culots de  $100 \times 10^6$  cellules des lignées JY, EL4 et A20.

## **Matériel**

- PBS : 311-010-CL, Multicell
- RPMI 1640 1X : 350-000-CL, Multicell
- DMEM : 319-005-CL, Multicell
- FBS : 98150, Multicell
- Pen/Strep : 450- 201-EL, Multicell
- Glutamine : 609-065-EL, Multicell
- Pyruvate de sodium (100mM) 100X : 600-110-EL, Multicell
- Solution Hepes 1M: 330-050-EL, Multicell
- D-Glucose : code 0188-500G, Amresco
- B-Mercaptoethanol : 190242, MP Biomedical
- X-vivo 15 : 04-744Q ou Immumax - IX : 001-061-CL, Lonza
- Compteur de cellules : invitrogen Countess II Automated Cell Counter : AMQAX1000
- Chambres de comptage de cellules : Countess II : C10283
- Colorant bleu Trypan (0,4%) : Ref : EB7-001, Nano Entek
- Flacon : Falcon 250 mL col incliné bouchon bleu ventilé traités pour culture de tissus: Ref : 353136
- Tubes coniques Falcon 50 ml en polypropylène : Ref 352070
- Tubes coniques Falcon 15 ml polypropylène haute clarté : Ref : 352096

**Horaire Optimisé pour la culture des lignées JY et EL4 pour recueillir 20x 100x10<sup>6</sup> cellules par semaine.**

**Jour 1 :**

**Ensemencement :**

- **JY:** 8x Falcon 250 mL Ref: 353136
  - 30 mL de milieu DMEM, 10% de FBS et 1% pen/strep.
    - Concentration cellulaire de départ : 0,6x10<sup>6</sup> cellules/mL
- **EL4 :** 4x Falcon 250 mL Ref : 353136
  - 30 mL de milieu X-vivo 15, 25% de DMEM, 10% de FBS 1% pen/strep
    - Concentration cellulaire de départ : 0,4x10<sup>6</sup> cellules/mL

**Jour 4:**

**Séparation :**

- **JY :** Concentration après 72 heures : 2.5x10<sup>6</sup> cellules/mL
  - Séparer en 25 Falcon 250 mL Ref : 353136 et ajuster à un volume final de 30 mL
- **EL4 :** Concentration après 72 heures : 3.0x10<sup>6</sup> cellules/mL
  - Séparer en 25 Falcon 250 mL Ref : 353136 et ajuster à un volume final de 30 mL

**Jour 6 :**

**Récolte et ensemencement :**

- **JY :** Concentration après 48 heures : 2.0x10<sup>6</sup> cellules/mL
  - Recueillir 10 culots de 100x10<sup>6</sup> cellules

- Ensemencer 22 Falcon 250 mL Ref : 353136 à  $0,75 \times 10^6$  cellules/mL
- **EL4** : Concentration après 48 heures :  $2.0 \times 10^6$  cellules/mL
  - Recueillir 10 culots de  $100 \times 10^6$  cellules
  - Ensemencer 22 Falcon 250 mL Ref : 353136 à  $0,4 \times 10^6$  cellules/mL

**Jour 8:**

**Récolte et ensemencement :**

- **JY** : Concentration après 48 heures :  $2.0 \times 10^6$  cellules/mL
  - Recueillir 10 culots de  $100 \times 10^6$  cellules
  - Ensemencer 8 Falcon 250 mL Ref : 353136 à  $0,6 \times 10^6$  cellules/mL
- **EL4** : Concentration après 48 heures :  $2.0 \times 10^6$  cellules/mL
  - Recueillir 10 culots de  $100 \times 10^6$  cellules
  - Ensemencer 4 Falcon 250 mL Ref : 353136 à  $0,4 \times 10^6$  cellules/mL

## Annexe 2 - Protocole pour la quantification des récepteurs MHC de surface par QIFIKIT<sup>R</sup>

Ce protocole est utilisé pour la quantification des molécules de surface MHC sur les différentes lignées cellulaires à l'aide du protocole QIFIKIT<sup>R</sup> d'immunofluorescence indirecte pour la cytométrie de flux.

### Matériel :

- Tubes FACS (BD Falcon)
- Tubes 1.5 mL (Eppendorf)
- Tube 15 mL (BD Falcon)
- PBS/FBS 1% : FACS Flow (Maintenir sur glace)
- Anticorps primaire IgG (ex: W6/32: BioXcell, InVivoMAb anti-human MHC Class 1 (HLA-A, HLA-B, HLA-C) #BE0079)
- Contrôle primaire d'isotype IgG (ex : Biolegend, Purified mouse IgG2b, k isotype ctrl, Cat : 402202)
- Anticorps secondaire marqué au fluorochrome (ex : Dako, Polyclonal Rabbit Anti-Mouse Immunoglobulins/FITC Code F0232)
- Flacons QIFIKIT<sup>®</sup> 1,2 et "Setup-Beads" (Dako, cat#K0078)
- Thermo Countess II ou III (Thermo Fisher : Compteur de cellules)
- Countess<sup>™</sup> Cell Counting Chamber Slides, Thermo Fisher: Catalog number: C10228

### Notes

*Attention* : Utilisez une salle de culture cellulaire de niveau de biosécurité 2.

*Note 1* : Gardez les cellules ainsi que le PBS/FBS1% sur glace (4°C) en tout temps.

*Note 2* : Toutes les étapes de centrifugation doivent être effectuées à 4°C.

*Note 3* : Attendre au moins une semaine avant d'effectuer des analyses sur des cellules qui ont été fraîchement sorties de la congélation.

## **Méthode**

### *En salle de Culture Cellulaire*

#### **A. Préparez les échantillons cellulaires.**

1. Effectuez un comptage cellulaire pour chaque lignée cellulaire avec l'appareil Countess. Bien mélanger avant le comptage. Mélangez 10 µL de bleu Trypan à 0,4 % + 10µl de suspension cellulaire et pipetez dans la lame de comptage. Enregistrez toutes les mesures sur une clé USB pour référence.
2. Calculez le volume de culture cellulaire nécessaire pour le nombre de cellules nécessaires pour l'expérimentation (Entre 500'000 et  $1 \times 10^6$  cellules).
3. Transférez le volume total requis pour tous les échantillons dans un tube Eppendorf 1.5 mL ou un tube Falcon de 15 mL. Identifiez les échantillons

#### **B. Lavage pour éliminer le milieu de culture**

1. Centrifugez a 1500 rpm x 5 min a 4°C
2. Jetez le surnageant et ajouter 1mL de FACS Flow froid.

#### **C. Préparez les solutions d'anticorps et contrôles iso-type pour le marquage primaire**

1. Dans un volume de 100 µL de FACS Flow, ajoutez la quantité suffisante d'anticorps primaires ou contrôle iso-type pour atteindre une condition de saturation (effectuez un titrage de l'anticorps à utiliser au préalable.)
2. Retirer le FACS Flow des tubes d'échantillons de cellules de l'étape 1.c. et pipetez le 100 µL de dilution d'anticorps de l'étape 3.a. dans ce même tube.
3. Vortexez légèrement.
4. Incubez 15 minutes à température pièce. (Pendant ce temps, préparez l'étape suivante)

#### **D. Préparez les billes QIFIKIT<sup>®</sup> :**

1. Vortexez et pipetez 100µl du flacon 1, du flacon 2 et des billes de réglages dans 3 tubes 1.5mL, 15mL ou FACS et identifiez les.

2. Traitez les tubes contenant des billes comme des échantillons pour le reste du protocole.

#### **E. Lavage pour éliminer l'excès d'anticorps primaires.**

1. Ajouter 3ml de FACS Flow froid à chaque tube, mélanger doucement.
2. Centrifugez à 1500 rpm x 5 min à 4°C
3. Jetez le surnageant.
4. Répétez les étapes 5.a.b.c (pendant les étapes de centrifugations, préparez l'étape suivante)

#### **F. Préparez la solution d'anticorps pour le marquage secondaire**

1. Dans un tube Eppendorf 1.5 mL, ajoutez 2µL d'anticorps secondaire FITC Dako et 98 µL de FACS Flow pour chaque échantillon cellulaire à analyser. Ex : 6 échantillons cellulaires = 12 µL d'anticorps et 588 µL de FACS Flow
2. Ajoutez 100 µL de cette solution à chaque tube suite à l'étape 5.d.
3. Vortexez doucement pour détruire le culot.
4. Incubez 15 minutes à température pièce, à l'abris de la lumière.

#### **G. Lavez pour éliminer l'excès d'anticorps secondaire**

1. Ajouter 3ml de FACS Flow froid à chaque tube, mélanger doucement.
2. Centrifugez à 1500 rpm x 5 min à 4°C
3. Jetez le surnageant.
4. Répétez les étapes 7.a.b.c
5. Remettre en suspension les cellules dans 250 µL de FACS Flow dans un tube FACS.
6. Conserver à 4°C jusqu'à ce que vous soyez prêt pour l'analyse en cytométrie.

#### **H. Analysez les cellules dans le FACS CANTO II.**

1. Fixez le nombre d'événements enregistrés à 10 000.
2. Sélectionnez l'amplification logarithmique pour le paramètre de fluorescence détectant le fluorochrome FITC.

#### **I. Établir la fenêtre d'analyse.**

1. S'assurer que les cellules négatives et les billes de réglage sont affichées à l'intérieur de la fenêtre d'analyse (en performant l'étape 9.b.)
2. Acquérir les données des billes de réglage et des contrôles négatifs.
3. Si la détection des contrôles négatif et des billes dépasse la fenêtre d'analyse, ajustez les paramètres de détection de fluorescence du fluorochrome FITC afin qu'elles le soient.

#### **J. Acquisition de données**

1. Acquérir les données de chaque tube FACS à analyser, rassembler ~10 000 événements par tube.
2. Sauvegarder l'expérience et exécuter un cycle de nettoyage sur le cytomètre de flux.

#### **K. Analyse des données**

1. Dans la première fenêtre d'analyse, isoler cellules vivantes et placez-les dans une 2ième fenêtre.
2. Dans cette deuxième fenêtre, rejetez les cellules qui semblent être en amas et isolez les cellules uniques dans une 3<sup>ème</sup> fenêtre d'analyse.
3. Obtenez le MFI pour chaque population de billes de calibration et chaque tube d'expérience.
4. Obtenir la courbe d'étalonnage par une transformation log<sub>10</sub> du MFI et de la capacité de liaison des anticorps (ABC: les valeurs spécifiques du lot incluses dans chaque kit), et tracer  $\log(\text{ABC}) = a \times \log(\text{MFI}) + b$ .
5. Utiliser l'équation de régression linéaire de la courbe d'étalonnage établie en 11.d. pour transformer la valeur du MFI de chaque échantillon en ABC.
6. Soustraire l'ABC du contrôle iso-type approprié de chaque échantillon pour obtenir la capacité spécifique de liaison des anticorps (SABC). Le SABC correspond au nombre de récepteurs HLA de chaque échantillon.
7. Effectuez les moyennes des répliques pour avoir une quantification finale.

## Annexe 3 – Protocole détaillé de l'isolation des peptides HLA par immuno-affinité

### Matériel

- 100 millions de cellules des lignées à analyser.
- Micro tube, 2 ml Protein LoBind de Eppendorf, cat # 022431002
- Tubes Falcon 15 ml, Corning, cat # 352096
- Embouts P-1000, P200 et P10 de Eppendorf, à faible liaison
- Sépharose activé par CNBr (GE Healthcare, # 17-0430-01)
- Rotateur
- Micro tube, 1.5 mL Protein LoBind de Eppendorf, cat # 022431081
- Micro tube, 2 mL Protein LoBind de Eppendorf, cat # 022431002
- Tubes Falcon 15 ml, Corning, cat # 352096
- Embouts P-1000, P200 et P10 de Eppendorf, à faible liaison (de Fisher : #2717352, 2717351, 2717349)
- Colonnes Bio-Rad - Colonnes de chromatographie Poly prep cat 731-1550
- Disque d'extraction en phase solide, colonne ultramicrospin C18 (The est group) Catalogue # SEMSS18V. Attention, toutes les colonnes ont le même numéro de produit mais une capacité volumique différente. Prenez celle de 5-200 µl.
- Bio-Rad Rack (cat# 191-2292)

### Notes

*Note 1* : Préparez de nouvelles solutions chaque fois que vous commencez une nouvelle expérimentation.

*Note 2* : Lors de l'utilisation d'un nouvel anticorps, nous suggérons de vérifier l'efficacité de la liaison en prélevant des aliquotes témoins (voir l'indication \*

Facultatif) et en effectuant une coloration au bleu de Coomassie sur gels SDS-PAGE.

### **Couplage d'anticorps à des billes activées au sépharose CNBr**

#### **Solution de réactivation pour le sépharose activé par CNBr : 1 mM HCL.**

- Pour 2 échantillons (15 ml par échantillon) :
- HCl 6N: 11,1  $\mu$ l + 50 ml H<sub>2</sub>O

#### **Tampon de couplage d'anticorps : NaCl 0,5 M, NaHCO<sub>3</sub> 0,1 M, pH 8,3.**

- 2,92 g NaCl
- 0,84 g NaHCO<sub>3</sub> (Amresco # 0865-1kg)
- Ajouter 100 ml d'eau
- Mélanger et ajuster le pH (16  $\mu$ l de NaOH 10 N)

#### **Solution de blocage pour le sépharose activé par CNBr : glycine 0,2 M.**

- 0,15 g de glycine
- Ajouter 10 ml de H<sub>2</sub>O

#### **A. Activation des billes de sépharose CNBr**

1. Peser 80 mg de billes CNBr par échantillon et les transférer dans un tube conique de 15 mL.
2. Remettre doucement en suspension les billes avec 5 mL de HCl 1 mM.
3. Ajouter 8,5 mL de HCl 1 mM aux billes.
4. Faire tourner lentement pendant 30 min à température ambiante.
5. Centrifuger les billes à 200 x g pendant 2 min puis éliminer le surnageant.
6. Ajouter 500  $\mu$ L de tampon de couplage au culot de billes.

#### **B. Couplage d'anticorps**

1. Transférer les billes dans un nouveau tube et ajouter 2 mg d'anticorps, puis ajuster le volume à 1000 $\mu$ L avec la solution tampon de couplage.

\*Optionnel : Prendre une aliquote de 18  $\mu$ L (INPUT), ajouter le tampon SDS-PAGE 4x et congeler immédiatement.

2. Faire tourner lentement pendant 120 min à température ambiante.
3. Centrifuger les billes à 200 x g pendant 2 min, puis retirer le surnageant.

\*Optionnel : Prendre une aliquote de 18  $\mu$ L du surnageant (anticorps non lié), ajouter 6  $\mu$ L de tampon SDS-PAGE 4x et congeler immédiatement.

### **C. Blocage et lavage des billes couplées à des anticorps**

1. Ajouter 1 mL de solution de blocage glycine 0.2 M.
2. Faire tourner lentement pendant 60 min à température ambiante.
3. Centrifuger les billes à 200 x g pendant 2 min puis éliminer le surnageant.
4. Ajouter 1mL de PBS.
5. Centrifuger les billes à 200 x g pendant 2 minutes puis éliminer le surnageant. \*Optionnel : Prendre une aliquote de 18  $\mu$ L du surnageant (anticorps non lié, dernier lavage), ajouter 6  $\mu$ L de tampon SDS-PAGE 4x et congeler immédiatement.
6. Ajouter 1mL de PBS. \*Optionnel : Prélever une aliquote de 18  $\mu$ L du mélange de billes (billes couplées à l'anticorps), ajouter 6  $\mu$ L de tampon SDS-PAGE 4x et congeler immédiatement.
7. Conserver à 4 degrés Celsius jusqu'à utilisation.

### **D. Lyse cellulaire et immuno-capture avec des billes couplées à un anticorps**

- **Tampon de lyse 2x : 1.0 % p/v CHAPS (EMDMillipore 220201-10MG)**
- 0.1g Chaps/10ml PBS pour 1.0%)
- Ajouter 1 pellet d'inhibiteur de protéines et de protéases Invitrogen (1 pellet = 0.04g)
- Conserver sur la glace
  1. Décongeler le culot cellulaire dans le tube en le réchauffant dans la paume de la main.
  2. Ajouter 500  $\mu$ L de PBS au culot et pipeter de haut en bas jusqu'à homogénéisation.

3. Mesurer le volume total et le transférer dans un nouveau tube.
4. Ajuster le volume du culot cellulaire à deux fois son volume original en utilisant 1% de Chaps dans du PBS contenant des inhibiteurs de protéase (jusqu'à une concentration finale de 0,5% de Chaps).
5. Tourner lentement pendant 60 min à 4°C
6. Centrifuger les cellules à 18000 x g pendant 20 min à 4°C
7. Utiliser le surnageant pour la procédure d'immunoprécipitation.
8. Centrifuger les billes à 200 x g pendant 2 min, puis retirer le surnageant.
9. Transférer le surnageant de lysat cellulaire sur l'anticorps couplé aux billes, puis incubé 14-18 heures avec une rotation lente à 4°C.

#### **E. Élué des complexes CMH-peptide à partir des billes**

- **Tampon C Tris 20 mM pH 8.0 (Tris-HCl, #10812846001)**
  - $157,6\text{g/mol} \times 20 \times 10^{-3} \text{ mol/1000 ml} \times 1000 \text{ ml} = 3,14 \text{ g Tris HCl dans 1L H}_2\text{O}$
  - pH 8.0 : (plus ou moins 650  $\mu\text{L}$  de NaOH 10N)
  - **Tampon A : 150 mM NaCl, 20 mM Tris, pH 8,0**
  - $58,44\text{g/mol} \times 150 \times 10^{-3} \text{ mol/1000ml} \times 600 \text{ ml} = 5.26 \text{ g de NaCl dans 600 ml de Buffer C}$
  - **Tampon B : 400 mM NaCl, 20 mM Tris, pH 8.0**
  - $58,44 \text{ g/mol} \times 400 \times 10^{-3} \text{ mol/1000 ml} \times 100 \text{ ml} = 4,68 \text{ g NaCl dans 200 ml de tampon C}$
  - **1% TFA**
  - 100  $\mu\text{L}$  TFA + 9,9 ml H<sub>2</sub>O pour 10 ml)
1. Placez les colonnes Bio-Rad sur le portoir Bio-Rad et placez un récipient vide en dessous pour recueillir le flux.
  2. Laver la colonne Bio-Rad avec 10 mL de tampon A et laisser s'écouler par gravité.

3. Récupérer le mélange billes-lysate et le transférer dans la colonne Bio-Rad.  
\*OPTIONNEL : Prélever une aliquote de 20µl (sur le volume total) pour le western blotting et congeler immédiatement.
4. Laver le tube contenant le mélange lysate-billes avec 1 mL de tampon A et transférer dans la colonne Bio-Rad pour récupérer les billes autant que possible.
5. Laver les billes par gravité avec 10 mL de tampon A.
6. Laver les billes par gravité avec 10 mL de tampon B.
7. Laver les billes par gravité avec 10 mL de tampon A.
8. Laver les billes par gravité avec 10 mL de tampon C.
9. Placer la colonne Bio-Rad dans un nouveau tube de 2.0 ml et élué les peptides liés au CMH en ajoutant 300 µL de TFA à 1%.
10. Mélanger les billes en pipetant de haut en bas 5 fois et laisser la solution s'élué et la transférer dans un nouveau tube.
11. Répétez une fois les étapes 9 et 10 et regrouper les élués. \* OPTIONNEL : Prélever une aliquote de 10µl pour le western blotting et congeler immédiatement.

#### **F. Dessalage et élution des peptides du CMH**

- **TFA 0,1%**
- 10 µL TFA + 9,99 mL H<sub>2</sub>O pour 10 mL
- **80% ACN/0.1% TFA**
- 8ml ACN + 10 µL TFA+1.99mL H<sub>2</sub>O pour 10mL
- **28% ACN/0.1% TFA**
- 2.8 ml ACN + 10 µL TFA +7.2 mL H<sub>2</sub>O pour 10mL

**Note :** Les étapes de dessalage et d'élution des peptides du CMH peuvent être réalisées en installant la colonne C18 sur un tube de 2.0 mL. Pour un meilleur ajustement, installez les anneaux (2) fournis par le fabricant entre la colonne C18 et le tube de 2.0 mL. Dans ce protocole, nous utilisons une colonne C18 avec une capacité de volume de 5-200µL (6-60µg). Toutes les étapes sont réalisées à température ambiante.

1. Ajouter 200µL de Méthanol sur la matrice C18, puis centrifuger à 1546 x g pendant 3min.
2. Jeter le flux.
3. Ajouter 200µL de 80%ACN/0.1%TFA sur la matrice C18, puis centrifuger à 1546 x g pendant 3 min.
4. Jeter le flux.
5. Ajouter 200µL de TFA à 0.1% sur la matrice C18, puis centrifuger à 1546 x g pendant 3 min.
6. Jeter le flux.
7. Charger 200µL de peptides élués sur la matrice C18, puis centrifuger à 1546 x g pendant 3 minutes. Répéter cette étape jusqu'à ce que le volume complet ait été chargé. Jeter le flux entre chaque chargement.
8. Laver l'échantillon avec 200µL de 0.1%TFA, puis centrifuger à 1546 x g pendant 3 min.
9. Transférer l'embout C18 dans un tube propre.
10. Éluer l'échantillon avec 200µL de 28%ACN/0.1%TFA, puis centrifuger à 1546 x g pendant 3 min. Répétez cette étape deux fois pour un volume total de 600 µL. \*OPTIONNEL : Recueillir une aliquote de 10µl pour le western blotting et congeler immédiatement.
11. Évaporer l'échantillon jusqu'à siccité en utilisant un speed-vac.
12. Garder congelé jusqu'à ce que les échantillons soient analysés par LC-MS/MS.

**Annexe 4 - "MhcVizPipe: A Quality Control Software for  
Rapid Assessment of Small- to Large- Scale  
Immuno-peptidome Data Sets"**

## Journal Pre-proof



MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- to Large-Scale Immunopeptidome Data Sets

Kevin A. Kovalchik, Qing Ma, Laura Wessling, Frederic Saab, Jérôme Despault, Peter Kubiniok, David J. Hamelin, Pouya Faridi, Chen Li, Anthony W. Purcell, Anne Jang, Eustache Paramithiotis, Marco Tognetti, Lukas Reiter, Roland Bruderer, Joël Lanoix, Éric Bonneil, Mathieu Courcelles, Pierre Thibault, Etienne Caron, Isabelle Sirois

PII: S1535-9476(21)00150-X

DOI: <https://doi.org/10.1016/j.mcpro.2021.100178>

Reference: MCPRO 100178

To appear in: *Molecular & Cellular Proteomics*

Received Date: 21 May 2021

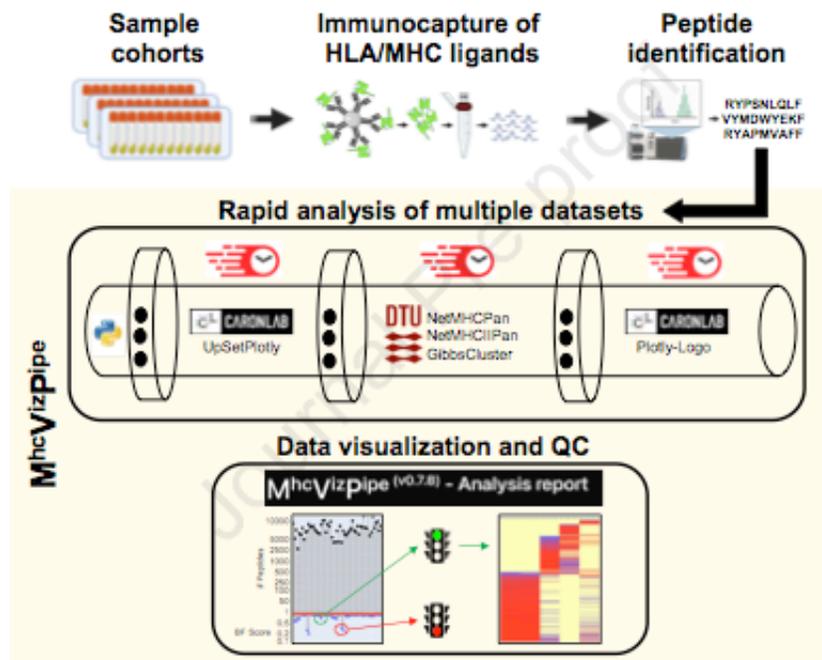
Revised Date: 28 October 2021

Accepted Date: 1 November 2021

Please cite this article as: Kovalchik KA, Ma Q, Wessling L, Saab F, Despault J, Kubiniok P, Hamelin DJ, Faridi P, Li C, Purcell AW, Jang A, Paramithiotis E, Tognetti M, Reiter L, Bruderer R, Lanoix J, Bonneil É, Courcelles M, Thibault P, Caron E, Sirois I, MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- to Large-Scale Immunopeptidome Data Sets, *Molecular & Cellular Proteomics* (2021), doi: <https://doi.org/10.1016/j.mcpro.2021.100178>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.



1 **MhcVizPipe: A Quality Control Software for Rapid Assessment of Small-**  
2 **to Large-Scale Immunopeptidome Data Sets**

3

4 **Authors**

5 Kevin A. Kovalchik<sup>1</sup>, Qing Ma<sup>2</sup>, Laura Wessling<sup>1</sup>, Frederic Saab<sup>1</sup>, Jérôme Despault<sup>1</sup>, Peter  
6 Kubiniok<sup>1</sup>, David J. Hamelin<sup>1</sup>, Pouya Faridi<sup>3</sup>, Chen Li<sup>3</sup>, Anthony W. Purcell<sup>3</sup>, Anne Jang<sup>4</sup>,  
7 Eustache Paramithiotis<sup>4</sup>, Marco Tognetti<sup>5</sup>, Lukas Reiter<sup>5</sup>, Roland Bröderer<sup>5</sup>, Joël Lanoix<sup>6</sup>, Éric  
8 Bonneil<sup>6</sup>, Mathieu Courcelles<sup>6</sup>, Pierre Thibault<sup>6,7</sup>, Etienne Caron<sup>1,8\*</sup>, Isabelle Sirois<sup>1\*</sup>

9

10 **Affiliations**

11 \*Corresponding authors: Etienne Caron ([etienne.caron@umontreal.ca](mailto:etienne.caron@umontreal.ca)) and Isabelle Sirois  
12 ([isabelle.sirois.phd@gmail.com](mailto:isabelle.sirois.phd@gmail.com))

13

14 <sup>1</sup>CHU Sainte-Justine Research Center, Montreal, QC H3T 1C5, Canada

15 <sup>2</sup>School of Electrical Engineering and Computer Science, Faculty of Engineering, University of  
16 Ottawa, ON K1N 6N5, Canada

17 <sup>3</sup>Infection and Immunity Program and Department of Biochemistry and Molecular Biology,  
18 Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

19 <sup>4</sup>CellCarta, Montreal, QC H2X 3Y7, Canada

20 <sup>5</sup>Biognosys, Wagistrasse 21, 8952 Schlieren, Switzerland

21 <sup>6</sup>Institute of Research in Immunology and Cancer, Montreal, QC H3T 1J4, Canada

22 <sup>7</sup>Department of Chemistry, Université de Montréal, Montreal, QC H3T 1J4, Canada

23 <sup>8</sup>Department of Pathology and Cellular Biology, Faculty of Medicine, Université de Montréal,  
24 QC H3T 1J4, Canada

25

26 **List of abbreviations**

27

28	BA:	Binding affinity
29	BF:	Binding fraction
30	DDA:	Data-dependent acquisition
31	DIA:	Data-independent acquisition
32	EL:	Eluted ligand
33	GUI:	Graphical user interface
34	HLA:	Human leukocyte antigen
35	LF:	Length fraction
36	MHC:	Major histocompatibility complex
37	MS:	Mass spectrometry
38	MVP:	MhcVizPipe
39	QA:	Quality assurance
40	QC:	Quality control

41

42

43

44

45

46

47

48

49 **ABSTRACT**

50 Mass spectrometry (MS)-based immunopeptidomics is maturing into an automatized, high-  
51 throughput technology, producing small- to large-scale datasets of clinically relevant MHC class  
52 I- and II-associated peptides. Consequently, the development of quality control (QC) and quality  
53 assurance (QA) systems capable of detecting sample and/or measurement issues is important for  
54 instrument operators and scientists in charge of downstream data interpretation. Here, we created  
55 MhcVizPipe (MVP), a semi-automated QC software tool that enables rapid and simultaneous  
56 assessment of multiple MHC class I and II immunopeptidomic datasets generated by MS, including  
57 datasets generated from large sample cohorts. In essence, MVP provides a rapid and consolidated  
58 view of sample quality, composition and MHC-specificity to greatly accelerate the 'pass-fail' QC  
59 decision-making process toward data interpretation. MVP parallelizes the use of well-established  
60 immunopeptidomic algorithms (NetMHCpan, NetMHCIIpan and GibbsCluster) and rapidly  
61 generates organized and easy-to-understand reports in HTML format. The reports are fully portable  
62 and can be viewed on any computer with a modern web browser. MVP is intuitive to use and will  
63 find utility in any specialized immunopeptidomic laboratory and proteomics core facility that  
64 provides immunopeptidomic services to the community.

65

66

67 **INTRODUCTION**

68 The importance of mass spectrometry (MS)-based immunopeptidomics for the discovery of T cell  
69 targets in autoimmunity (1, 2), cancer (3–5) and infectious diseases (6–10)—including pandemic  
70 pathogens (11, 12)—has attracted the interest of investigators from a wide range of clinical  
71 disciplines, leading to the creation of the human immunopeptidome project (13–15). In fact, the  
72 growing interest for clinical immunopeptidomics was recently accelerated by the unquestionable  
73 contribution to the 2018 Nobel prize winners James P. Allison and Tasuku Honjo for their work  
74 on cancer immunotherapy (16). In addition, the development of next-generation MS technologies  
75 and methods has fostered the discovery of new classes of actionable tumor-specific antigens (TSA)  
76 in various cancer types (17–22). In its simplest form, clinical immunopeptidomics involves the  
77 isolation of HLA class I- and class II-associated peptides from patient biospecimens by  
78 immunoaffinity capture, followed by peptide release, and subsequent peptide sequence  
79 identification by MS combined with advanced bioinformatics (23). Once the HLA ligands have  
80 been confidently identified, their immunogenicity can be evaluated to further guide the  
81 development of vaccines and T-cell based therapies in translational laboratories (24).

82 Data quality is a cornerstone of solid research, demanding repeatability and reproducibility  
83 (25). In that respect, assessing and controlling the quality of immunopeptidomic data generated by  
84 MS is of utmost importance. In genomics and MS-based proteomics, the importance of quality  
85 control (QC) and quality assurance (QA) has been long acknowledged, and various grades of QC  
86 samples (i.e. QC1, QC2 and QC3) (26) as well as QC software solutions have been extensively  
87 developed and applied over the years (27–36). In contrast, in MS-based immunopeptidomics, QC  
88 samples and QC software tools specialized for MHC class I- and II-associated peptides remain  
89 poorly documented in spite of their importance for successful therapeutic development (37).

90 To date, only two studies have focused on QC measures to validate the quality of  
91 immunopeptidomic data generated by MS (37, 38). In the first study, Ghosh et al. (2020) describes  
92 the importance of i) the mass accuracy of the obtained peptides, ii) the fitness of the observed and  
93 calculated retention times, iii) the repeatability of the retention times and the signal intensities of  
94 the detected peptides, iv) the use of synthetic peptides (light or heavy) to determine the specificity  
95 and the limit of detection of the HLA ligands of interest, and v) the technical and biological  
96 reproducibility (37). The identification scores of the peptides and their identification with different  
97 search engines can also be applied to assess data quality (39). In the second QC study, Fritsche et  
98 al. (2021) presents i) statistics to enable discrimination of true HLA ligands from coisolated HLA-  
99 independent proteolytic fragments, ii) the necessary steps to ensure system suitability of the  
100 chromatographic system, iii) an algorithm for detection of source fragmentations events that are  
101 introduced by electrospray ionization during MS, and iv) an experimental pipeline that enables  
102 high-throughput sequence verification through similarity of fragmentation patterns and coelution  
103 of synthetic isotope-labeled internal standards (38). Akin to MS-based proteomics, such QC  
104 approaches are useful to show the overall quality of immunopeptidomic datasets in addition to  
105 point out limitations and pitfalls that are critical for individual peptides.

106 Other basic and essential steps to assess the overall quality and MHC-specificity of  
107 immunopeptidomic datasets are to quantify the total number of peptides per sample, the length of  
108 the detected peptides, the number of strong, weak and non-MHC binders per sample, the number  
109 of strong, weak and non-MHC binders per allele per sample, the number of peptides making up  
110 each sequence motif and the fraction of each sequence motif attributed to each MHC allele. Current  
111 software tools used to assess such QC measures to determine the MHC-specificity of  
112 immunopeptidomic datasets include MHC peptide binding prediction algorithms and clustering

**Running title: MhcVizPipe software**

113 tools such as NetMHCpan (40, 41), MHCFlurry (42, 43), GibbCluster (44) and MoDec (45).  
114 Although widely used, these algorithms were not purposely built for QC in MS-based  
115 immunopeptidomics, and as a result, can process only one sample at a time and generally require  
116 further human-based data manipulations (e.g. in Excel)—a relatively time-intensive and error-  
117 prone procedure that is not sustainable for QC in large-scale MS-based immunopeptidomics studies,  
118 as recently reported (46, 47). Hence, the development of automated or semi-automated QC  
119 software tools for rapid and simultaneous quality assessment of the MHC-specificity of multiple  
120 immunopeptidomic datasets generated by MS has yet to be developed.

121 In this technical report, we document QC in MS-based immunopeptidomics by presenting  
122 MhcVizPipe (MVP). MVP is an open-source, freely available QC software tool with an intuitive  
123 graphical user interface (GUI), intended to be used by any immunopeptidomic laboratory. MVP  
124 builds upon the algorithms mentioned above (NetMHC suite tools and GibbsCluster) and, once  
125 installed, provides a semi-automated and fast post-data acquisition system to assess and control  
126 sample quality and MHC-specificity through effective visualization of one or multiple  
127 immunopeptidomic datasets in a HTML report (**Figure 1A**). Reported QC metrics are peptide  
128 length, numbers of MHC binders, distribution of binders per MHC allele and prominent MHC  
129 peptide binding motifs (see examples of HTML reports in **Supplemental Data S1-S4**). In addition,  
130 MVP computes length fraction (LF) scores and binding fraction (BF) scores for each sample to  
131 rapidly highlight problematic samples while analyzing large immunopeptidomic sample cohorts.  
132 In this regard, a command line interface is also available for large-scale / 'batch' analyses  
133 (<https://github.com/CaronLab/MhcVizPipe/wiki/Command-line-interface>). Below, we describe i)  
134 the installation procedure of MVP, ii) the MVP graphical user interface for data upload, iii) the  
135 multiple computational steps that are automatically performed within the MVP-HTML reporting

**Running title: Mincvizrpe software**

136 pipeline, iv) the content of an HTML report, v) QC analysis of low- to high-quality biopsies, vi)  
137 the speed performance of MVP, and vii) QC analysis of a large cohort composed of 152 samples.  
138 We also discuss the current limitations of MVP and how the software could be further developed  
139 to support QC in large-scale clinical immunopeptidomic studies.

140

**141 METHODS****142 System requirements**

143 *Operating system.* MVP runs natively on Linux (e.g. Ubuntu) or MacOS systems. It can also be  
144 installed on Windows 10 systems by using the Windows Subsystem for Linux (WSL) to run  
145 NetMHCpan, NetMHCIIpan and GibbsCluster. MVP was tested on the following systems: Ubuntu  
146 16.04, 18.04 and 20.04, Linux Mint 18, MacOS 10.13 HighSierra, 10.14 Mojave, 10.15 Catalina,  
147 11.2.3 Big Sur and Windows 10 with both WSL 1 and WSL 2. *Memory and processor.* There are  
148 no hard memory or CPU requirements for MVP. However, because MVP utilizes multi-threading,  
149 performance increases on systems with higher numbers of CPUs. The memory usage is minimal  
150 and should not present an issue to any recent desktop or laptop computer.

151

**152 Components of MVP**

153 MVP connects the bioinformatics tools NetMHCpan, NetMHCIIpan, and GibbsCluster with  
154 Python visualization libraries to create portable HTML reports for interpretation of  
155 immunopeptidomics MS data. In addition to NetMHCpan, NetMHCIIpan, and GibbsCluster, it  
156 makes use of the following third-party Python libraries: Plotly, PlotlyDash, Dash Bootstrap  
157 Components, Pandas, Numpy, Dominate, UpsetPlotly, Waitress and PlotlyLogo (PlotlyLogo and

7

**Running title: MHCviz peptide software**

158 UpSetPlotly developed in-house and describe below). The graphical user interface of MVP is built  
159 using the PlotlyDash library and runs as a local web application (i.e. it runs in a web browser).

160

**161 Peptide list pre-processing**

162 Prior to analysis, peptide lists are stripped of chemical modifications (e.g. oxidation of methionine  
163 and carbamidomethylation), any peptides containing non-standard amino acids are removed, and  
164 flanking amino acids are removed if found (e.g. P and V in P.KAPDNRETL.V). It is assumed  
165 flanking amino acids are separated from the main sequence with dots, as indicated in the example.  
166 Formats differing from this will require preprocessing by the user.

167

**168 Length distributions and intersecting peptide sets**

169 Length distributions are visualized for all peptides  $\leq 30$ -mer in the input. If multiple samples are  
170 analyzed, the intersections of the peptide sequence sets arising from the samples are visualized in  
171 an UpSetPlot-type figure (48) using the Python library UpSetPlotly.

172

**173 NetMHCpan and NetMHCIIpan**

174 Prior to analysis with NetMHCpan or NetMHCIIpan, subsets of the peptide lists are created as  
175 follows: for class I, the lengths are restricted to between 8 and 12 mer inclusive; for class II, the  
176 lengths are restricted to between 9 and 22 mer inclusive. These lists are analyzed using either  
177 NetMHCpan or NetMHCIIpan to yield binding predictions (eluted ligand percent rank) used to  
178 annotate the peptides as follows: for class I, strong binders rank is  $\leq 0.5$ , weak binders rank is  $\leq$   
179 2.0 and non-binders rank is  $> 2.0$ ; for class II, strong binders rank is  $\leq 2.0$ , weak binders rank is  $\leq$   
180 10 and non-binders rank is  $> 10$ . The results are presented in tabular format as well as with a bar

**Running title: Mincvizipe software**

181 plot and heat map. The bar plot shows the number of peptides versus aggregate binding strength  
182 (e.g. a peptide that has a weak affinity for one allele and a strong affinity for another is counted as  
183 a strong binder). The heatmap is sorted by percent rank eluted ligand score from left-to-right (i.e.  
184 the peptides are ordered first by the left-most column, then the second-left-most, etc.). To prevent  
185 the convolution of the ordering by poorly scoring peptides, all values greater than 2.5 for class I or  
186 12 for class II are set to 2.5 or 12, respectively. The colormap of the heatmap is set such that red  
187 approximately represents strong binders, blue approximately represents weak binders, and yellow  
188 represents non binders with a linear gradient between the colors.

189

**190 Length fraction and binding fraction score**

191 The length fraction (LF) score is the fraction of all peptides which are 8-14 mers in length for class  
192 I peptides, or 8-25 mers in length for class II peptides. The binding fraction (BF) score is the  
193 fraction of all peptides within the above length range which are predicted to be either weak or  
194 strong binders by NetMHCpan or NetMHCIIpan.

195

**196 GibbsCluster**

197 MVP performs two GibbsCluster routines which we have termed "Unsupervised GibbsCluster"  
198 and "Allele-Specific GibbsCluster". There is a tab for each of these in the "Sequence Motifs"  
199 section of the report. The unsupervised GibbsCluster is a standard GibbsCluster run using all  
200 peptides in the subset described above. The following parameters are used, which are the  
201 recommended defaults for class I and II peptides on the GibbsCluster-2.0 server  
202 (<https://services.healthtech.dtu.dk/service.php?GibbsCluster-2.0>): -g 1-6 -T -j 2 -C -D 4 -I 1 (class  
203 I); -g 1-6 -k 1 -T -j 2 (class II). Note that the grouping setting (-g) is actually run one group at a

**Running title: MHCvizipe software**

204 time, as explained in the following section. The grouping with the highest KLD score is presented  
205 in the report. The “Allele-specific GibbsCluster” is dependent upon the results of NetMHCpan or  
206 NetMHCIIpan. A subset of peptides is created for each allele, such that all the peptides in a subset  
207 are strong or weak binders for the respective allele. An additional subset is created in which  
208 peptides predicted to be non-binders for all the alleles are combined (i.e. the peptides in this set are  
209 not predicted binders for any allele). Any subset containing less than 20 peptides is discarded. Each  
210 of the remaining subsets is run in GibbsCluster using the above parameters with -g set to 1, forcing  
211 GibbsCluster to look for only one peptide group. An exception is the subset of non-binders, in  
212 which -g is set to 1-5, allowing GibbsCluster to look for multiple groups in these unannotated  
213 peptides. As above, the grouping with the highest KLD score is presented in the report. The results  
214 are reported in the “Allele-Specific GibbsCluster” tab, where we see one motif for each allele  
215 present in the sample, and up to 5 motifs for the non-binding peptides. All the peptide groups shown  
216 in the “Sequence Motifs” section are visualized using the PlotlyLogo library as described below.

217

**218 Multiprocessing in NetMHCpan, NetMHCIIpan and GibbsCluster**

219 To shorten the overall analysis time, MVP parallelizes the use of NetMHCpan, NetMHCIIpan and  
220 GibbsCluster. Because NetMHCpan and NetMHCIIpan do not take advantage of multiprocessing,  
221 in MVP the peptide lists are broken into smaller lists which are concurrently analyzed by separate  
222 instances of the respective software. The results are then combined back into a single list.  
223 GibbsCluster does offer multiprocessing, analyzing the different grouping possibilities (e.g. 1, 2,  
224 3, 4, or 5, groups) concurrently. Unfortunately, with larger peptide lists, the single-group clustering  
225 is often long-running, which makes it difficult to schedule consecutive GibbsCluster analyses in an  
226 efficient manner. To address this, MVP splits the analysis into individual instances of GibbsCluster,

**Running time: nvcvizirpe software**

227 each analyzing only a single grouping possibility. For example, instead of using the command line  
228 parameter “-g 1-5” as indicated in the previous sections, in actuality MVP creates five jobs with  
229 the parameters -g 1, -g 2, -g 3, etc. With a single sample this doesn't result in a speed improvement,  
230 but with multiple samples MVP is able to efficiently schedule the jobs from multiple samples to  
231 utilize all available processors.

232

**233 UpSetPlotly**

234 UpSetPlotly is an open-source Python package, based upon UpSet (48), developed in-house for  
235 visualizing intersecting sets using the Plotly Python library. The source code is available at  
236 <https://github.com/kevinkovalchik/UpSetPlotly> and it can be installed as 'upsetplotly' from PyPI  
237 using the Pip package manager.

238

**239 PlotlyLogo**

240 PlotlyLogo is an open-source Python package developed in-house for generating sequence logos  
241 from sequence alignment data and is available at <https://github.com/kevinkovalchik/Plotly-Logo>.  
242 PlotlyLogo was designed for the specific purpose of generating sequence logos from sequence  
243 alignments as native Python objects using the Plotly plotting framework. As such, the version used  
244 in this report (plotly-logo v0.0.2) does not include much of the functionality of more complete  
245 solutions such as Seq2Logo, but it has the advantage of utilizing a modern Python framework  
246 (compatible with Python3) and of generating figures as Python objects, which can be directly used  
247 in other Python code. The choice to develop PlotlyLogo rather than using an existing Python  
248 solution such as LogoMaker (49) was influenced by the desire to generate live, interactive figures  
249 in a portable HTML format. It is developed in the Python programming language and requires the

**Running title: MHCvizRipe software**

250 Python plotting library Plotly. It is installable as 'plotly-logo' from PyPI using the Pip package  
251 manager. The algorithms in PlotlyLogo are based upon the methods described by Thomsen et al.  
252 for Seq2Logo (50) and in Immunological Bioinformatics (51). In brief, sequence alignments are  
253 read from text-formatted files and probability matrices are generated using sequence weighting  
254 from Hobohm 1 clustering and pseudo count correction, as described (50, 51). Two types of  
255 sequence logos can be generated: Shannon and Kullback-Leibler.

256

**257 Datasets**

258 To test the performance of EL versus BA in annotating peptide for specific MHC alleles, the  
259 following datasets were used: H2-K<sup>b</sup> and H2-D<sup>b</sup> class I peptides extracted from mouse liver tissue  
260 (PXD008733) (52), HLA-A and -B class I peptides extracted from PBMC (PXD001872) (39), H2-  
261 IA<sup>d</sup> and H2-IE<sup>d</sup> class II peptides extracted from the mouse A20 cell line (53), and HLA-DQA10101  
262 -DQB10501 and HLA0DQA10103, -DQB10603 class II peptides extracted from the human  
263 MAVER-1 cell line (54). To test the speed performance of MVP, the following datasets were used:  
264 HLA-ABC-associated peptides extracted from the JY cell line and from multiple PBMC samples  
265 (PXD001872) (39), and H2-K<sup>b</sup>/H2-D<sup>b</sup> class I peptides extracted from various mouse tissues  
266 (PXD008733) (52). Supplementary materials in Rijensky et al. (55) and Shraibman et al. (46) was  
267 used to create Table 1, Figure 3 and Figure 4.

268

**269 Time speed estimation procedure**

270 To benchmark the speed performance of MVP, the selected datasets were processed manually  
271 (human-based) or using MVP (computer-based). Time estimates were measured using a stopwatch.  
272 For the computer-based approach, the data were processed with MVP using an iMac18,2 (MacOS

**Running title: MhcVizPipe software**

273 Big Sur Version 11.2.3) with 4 cores and 16GB memory. For the human-based approach, the exact  
274 same datasets were processed manually by an experienced researcher with expertise in  
275 immunopeptidomics, with the goal of generating figures and tables that are produced in a HTML  
276 report. In other words, the human-based approach mimicked as much as possible the computer-  
277 based approach to produce the equivalent of an HTML report. Hence, for the human-based  
278 approach, time estimates were measured using a stopwatch for the following actions: 1) make  
279 graphs and tables for peptide length distribution and specificity in Microsoft Excel, 2) run  
280 NetMHCpan 4.1 and GibbsCluster online for each dataset, 3) extract output files and manipulate  
281 the data manually for making histograms and heatmaps in Microsoft Excel from the predicted MHC  
282 peptide binding affinity scores, and 4) combine the results in a document.

283

**284 Batch analysis**

285 To facilitate automation or batch analysis—which might benefit from running MVP in a bash script,  
286 a python script, or as a scheduled process—MVP contains a command line interface (CLI). The  
287 CLI runs MVP and saves the HTML report, NetMHCpan predictions, and PDF files of all figures  
288 in the report in a specified location. The use of the CLI is described at the following link:  
289 <https://github.com/CaronLab/MhcVizPipe/wiki/Command-line-interface>.

290

**291 Code availability**

292 The MVP source code is open source and freely available at:  
293 <https://github.com/CaronLab/MhcVizPipe>. The version of MVP used in this publication is 0.7.9  
294 (<https://github.com/CaronLab/MhcVizPipe/releases/tag/v0.7.9>).

295

296 **RESULTS**297 **Installation of MVP**

298 MVP is freely available at <https://github.com/CaronLab/MhcVizPipe> and can be downloaded and  
299 installed on Linux (e.g. Ubuntu), Mac, and Windows 10 (using the Windows Subsystem for Linux).  
300 A detailed explanation of the installation process can be found on the MVP GitHub wiki:  
301 <https://github.com/CaronLab/MhcVizPipe/wiki>. Note that for installing and running MVP on  
302 Windows 10, a few additional steps are required (see  
303 <https://github.com/CaronLab/MhcVizPipe/wiki/Windows-installation> for details). In brief, two  
304 installation options are available: 1) a ZIP file containing a standalone Python distribution and all  
305 the Python packages required by MVP can be downloaded from the MVP GitHub repository  
306 (<https://github.com/CaronLab/MhcVizPipe/releases>), and 2) MVP can be installed from the Python  
307 Package Index into an existing Python environment. Both options also require the separate  
308 acquisition of copies of NetMHCpan-4.1, NetMHCIIpan-4.0 and GibbsCluster-2.0 from DTU  
309 Health Tech (<https://services.healthtech.dtu.dk/software.php>). Note that when new versions of  
310 NetMHCpan and GibbsCluster will be made publicly available, users will need to download the  
311 new versions themselves as there is no possibility at this time to automatically update the tools.  
312 When new versions are available, we will ensure that MVP functions with them.

313 Option 1 consists of a ZIP file which contains MVP, a standalone Python distribution, and  
314 all the required Python packages. The user must then copy or move the extracted DTU Health Tech  
315 tools (NetMHCpan, NetMHCIIpan and GibbsCluster) into the "tools" folder found inside the  
316 MhcVizPipe directory (**Figure 1B**). MVP can then be run by double-clicking the MhcVizPipe  
317 executable file or running the accompanying MhcVizPipe.sh script file. This installation option  
318 does not require any use of the terminal and the final program is fully portable (e.g. you could put

**Running title: MhcVizPipe software**

319 it on a USB drive and use it on any compatible computer, though use in Windows requires you to  
320 copy MVP from the USB to the hard drive before use).

321 While we have tried to ensure the portable installation is compatible with most systems, we  
322 cannot guarantee it will work on every computer. This stems from the fact that the standalone  
323 distribution of Python has system dependencies that are usually, but not always, present. To address  
324 this issue, it is also possible to install MhcVizPipe into an existing Python environment as described  
325 in "option 2".

326 Option 2 is intended for users who are familiar with the command line and Python and wish  
327 to install the MVP Python package themselves, or who are unable to use the portable distribution  
328 for the reasons mentioned above. MVP can be installed in an existing Python environment ( $\geq 3.7$ )  
329 from the Python Package Index by invoking "pip install MhcVizPipe" from the terminal. As above,  
330 the user must also acquire copies of NetMHCpan-(4.0 or 4.1), NetMHCIIpan-4.0 and  
331 GibbsCluster-2.0. For option 2, the user needs to ensure the tools from DTU Health Tech are  
332 properly configured (according to their accompanying documentation) and reference their  
333 locations in the MVP settings window after starting the program for the first time. With this type  
334 installation, the program is started from a terminal using this command: "python -m  
335 MhcVizPipe.gui".

336 Because MVP is run as a web server, it can be accessed by any computer connected to the  
337 local network, facilitating use by multiple users. If unexpected issues are encountered during the  
338 installation procedure, please contact the authors or open an issue at the following link:  
339 <https://github.com/CaronLab/MhcVizPipe/issues>. The proper installation of MVP can be tested  
340 using example peptide lists and accompanying HTML reports available at the following link:  
341 [https://github.com/CaronLab/MhcVizPipe/tree/master/test\\_data](https://github.com/CaronLab/MhcVizPipe/tree/master/test_data).

342

**343 The MVP graphical user interface**

344 MVP provides a simple and intuitive graphical user interface (GUI) in any web browser (tested in  
345 Firefox, Safari, Chrome, Chromium and Edge) (**Figure 2**). Peptide lists can be uploaded in different  
346 formats (.csv, .tsv or .txt) or copy-pasted into the GUI. Any file of the mentioned formats may be  
347 loaded provided they have columns headers or are a simple list. If a multi-column file is opened  
348 (e.g. database search results in .csv format), the user is asked by MVP to select the header of the  
349 column which contains the peptide sequences. There is no limit to the number of files which can  
350 be analyzed at one time, though processing time increases with the number of files and the  
351 interpretability of figures and tables will decrease with excessive numbers of samples. Once the  
352 peptide lists have been uploaded, the MHC class (I or II), and alleles corresponding to the sample(s)  
353 are specified. Up to six alleles may be specified for each sample. Many technical details related to  
354 the samples can also be specified and will appear in the final report, which is generated in a portable  
355 HTML format and can be viewed in the browser or saved to the computer. MVP also generates zip  
356 files containing editable PDF versions of all plots in the report and all the data used to generate the  
357 report (i.e. NetMHCpan predictions, GibbsCluster output files).

358

**359 Automated steps performed within the MVP reporting pipeline**

360 Peptide lists are stripped of peptides containing chemical modifications. The complete lists  
361 are then used for generating the length histogram and UpSet plot in the "Sample Overview" section  
362 of the HTML report. The lists are then subset to peptides 8-12 or 9-22 mers in length for class I or  
363 class II peptides, respectively. Binding predictions for these subsets are made using  
364 NetMHCpan4.0 or 4.1 or NetMHCIpan4.0 for each user-indicated allele and the eluted ligand

16

**Running title: MHCvizripe software**

365 percent rank scores are used to generate the “Annotation Results” and “Binding Heatmaps”  
366 sections in the HTML report. Length fraction (LF) and binding fraction (BF) scores are also  
367 calculated to provide numerical values regarding the overall quality of the samples. LF score is the  
368 fraction of all peptides which are 8-12 mers in length for class I peptides, or 9-22 mers in length  
369 for class II peptides; BF score is the fraction of all peptides within the appropriate length range  
370 which are predicted to be either weak or strong binders by NetMHCpan or NetMHCIIpan (**Table**  
371 **1**). Peptide grouping and alignment is then performed by GibbsCluster twice, once using the  
372 complete subset of peptides (yielding the results in the “Unsupervised GibbsCluster” tab) and then  
373 again for the sets of peptides predicted to bind the different alleles (yielding the results in the  
374 “Allele-Specific GibbsCluster” tab). Logos of the most prominent motifs identified by  
375 GibbsCluster (according to KLD scores) are generated using Plotly-Logo. Note that the goal of this  
376 clustering approach is not to provide new biological insights or highly accurate annotations of  
377 peptide subgroups to their alleles, but to rather provide an overall assessment of the quality and  
378 MHC-specificity of the data.

379

**380 MVP annotates peptides using the Elution Ligand (EL) method**

381 NetMHCpan (v4.1) and NetMHCIIpan (v4.0) are critical components of the MVP software for  
382 scoring class I and II peptides, respectively (**Figure 1A**). The predicted MHC binding affinity  
383 computed by these algorithms defines strong binders ( $\%Rank < 0.5$ ), weak binders ( $2.0 < \%Rank >$   
384  $0.5$ ) or non-binders ( $\%Rank > 2.0$ ) in the immunopeptidome dataset. Those predicted values are  
385 used to calculate the BF scores, which serve as an index of MHC-specificity for the samples.

386 The methods that are available for peptide scoring in NetMHCpan and NetMHCIIpan for  
387 peptide scoring are Eluted Ligand (EL) and Binding Affinity (BA) (41). Here, we compared both

**Running title: Mincvizipe software**

388 methods using four publicly available immunopeptidomic datasets (see Methods) to rationalize  
389 which method to apply and integrate as part of the MVP software. For class I peptides, we found  
390 that EL and BA performed equally well (**Supplemental Figure 1**). Indeed, the two methods  
391 resulted in very similar numbers of strong and weak binders, both in mouse (**Supplementary**  
392 **Figure 1A, B**) and human (**Supplementary Figure 1C, D**). In contrast, for class II peptides,  
393 substantial differences were observed between EL and BA, in both mouse (**Supplementary Figure**  
394 **2A, B**) and human (**Supplementary Figure 2C, D**). Specifically, we found that EL resulted in a  
395 nearly 4-fold increase in annotated peptides compared to BA. Furthermore, we found that the  
396 peptide groups identified by GibbsCluster contained a much higher proportion of strong binders  
397 when using EL (up to 75%) versus BA (up to 14%) (**Supplementary Figure 2B, D**). These results  
398 are in agreement with a recent study indicating that prediction scores based on EL were more  
399 accurate than those using BA, for class II peptides in particular (56). Thus, the EL method was  
400 chosen and implemented in MVP.

401

**402 Content of the HTML report**

403 The HTML report generated from the MVP graphic user interface contains three main sections: 1)  
404 Sample Overview, 2) Annotation Results and Binding Heatmaps, and 3) Sequence Motifs.  
405 Examples of HTML reports are provided in **Supplemental Data S1** and **S2** for mouse and human  
406 class I peptides, respectively; and in **Supplemental Data S3** and **S4** for mouse and human class  
407 II peptides, respectively.

408 The first section of the HTML report (Sample Overview) contains i) a peptide length  
409 distribution graph (up to 30-mers), ii) a descriptive table indicating the total number of peptides  
410 per sample, the number of peptides corresponding to the expected length related to the MHC class

**Running title: MHCvizRipe software**

411 selected, and LF and BF scores for each sample, and iv) an "UpSet" plot (48) showing the number  
412 of unique and shared peptides between multiple samples.

413 In the second section of the HTML report (Annotation Results and Binding Heatmaps),  
414 MVP uses the EL scoring and annotation method described above to generate histograms and  
415 heatmaps illustrating the proportion of immunoaffinity purified peptides predicted to bind MHC  
416 molecules. Those graphs provide a bird's eye view on the quality and MHC-specificity of the  
417 samples and is a way visualize the BS scores calculated in the 'Sample Overview' section.

418 The third section of the HTML report focuses on the identification and visualization of  
419 sequence motifs that are enriched in the analyzed samples using GibbsCluster and our in-house  
420 Plotly-Logo algorithm, respectively (see Methods). The sequence motifs from the GibbsCluster  
421 analysis are presented in two ways: i) the unsupervised tab displays the most prominent motif(s)  
422 represented by the subset of peptides in the sample along with the % of peptides associated with  
423 each allele, and ii) the allele-specific cluster tab displays peptide motif(s) generated from strong  
424 and weak MHC binders for each allele, as well as motifs from peptides predicted to be non-assigned  
425 MHC binders. Together, MVP generates in a few clicks a complete HTML report for assessing the  
426 quality and MHC-specificity of immunopeptidomic data generated by MS.

427

**428 QC analysis of tumor biopsies**

429 The general quality and MHC-specificity of immunopeptidomic data is assessed using LF and BF  
430 scores, as mentioned above. Such numerical values are useful to rapidly detect abnormalities to  
431 determine if samples are of high, middle or low quality. To show the utility of this scoring approach,  
432 we selected a small immunopeptidomic dataset generated from eight representative tumor biopsies,  
433 recently published in Rijensky et al. (55). This dataset is composed of biopsies of different cancer

434 types, including head and neck, bile duct, lung, gastric, vascular, bladder and pancreatic cancer.  
435 Sample weight of those biopsies ranged from < 13 mg to 150 mg, and number of peptides detected  
436 ranged from 144 peptides (bile duct; cholangiocarcinoma) to 1018 peptides (head and neck;  
437 squamous cell carcinoma) (**Table 1**). To calculate the LF scores, the number of peptides between  
438 8 and 12 mers per biopsy sample were divided by the total number of peptides detected. To  
439 calculate the BF score, the number of peptides predicted to be SB or WB (NetMHCpan %Rank)  
440 were divided by the number of peptides between 8-12 mers. Our data show that the LF scores  
441 varied from 0.43 to 0.95 and the BF scores varied from 0.06 to 0.98 (**Table 1**)—a numerical value  
442 of 1 being the highest score. Overall, those samples were qualified of low to high quality. For  
443 instance, biopsies 5, 6 and 7 yielded high quality immunopeptidomic data, with an average  
444 combined score (LF and BF) of 0.91. Indeed, the proportion of peptides between 8-12 mers that  
445 were predicted to bind the HLAs expressed in these biopsy samples were relatively high, as  
446 visualized in **Figure 3A**. In contrast, biopsy 2 yielded low quality immunopeptidomic data, with  
447 an average combined score (LF and BF) of 0.27 (**Table 1**). This relatively low score can be  
448 explained by the mere absence of 8-12 mers predicted as SB or WB (**Figure 3C**). Finally, biopsies  
449 3 and 4 yielded middle quality immunopeptidomic data, with an average combined score (LF and  
450 BF) of 0.42 (**Table 1**). This score can be explained by the noticeable presence of SB and WB within  
451 a relatively large proportion of peptides that were observed to be longer than 12 mers (likely  
452 contaminant peptides), and therefore, not predicted to bind the HLAs expressed in these biopsy  
453 samples (**Figure 3B**). Thus, our analysis show that LF and BF is a simple QC scoring approach  
454 that is integrated within the MVP software tool to determine the overall quality and MHC-  
455 specificity of HLA peptidomes isolated from human biopsies of various cancer types.

456

**457 High-speed performance of MVP**

458 A major advantage of processing immunopeptidomic data through MVP is the speed at which  
459 HTML reports are generated for one or multiple samples. To benchmark the speed performance of  
460 MVP, we measured the precise time estimates to assess data quality of five different datasets using  
461 i) the human/manual-based approach, as currently done in many laboratories, or ii) the  
462 computer/MVP-based approach (see Methods) (**Figure 4A**). Briefly, the commonly applied  
463 human-based approach includes uploading peptide lists and running multiple web-interfaces,  
464 downloading the results, editing them in Excel, making multiple figures and then copying them  
465 into a reporting document. The computer/MVP-based approach includes uploading peptide lists on  
466 the GUI, selecting the appropriate alleles, and clicking the 'GO!' button (**Figure 2**). The datasets  
467 that were selected to benchmark the software included a range of different samples per dataset  
468 (**Figure 4A**). By comparing the two approaches, our results show that the measured time estimates  
469 varied from ~20 minutes (for 1 sample) to ~400 minutes (for 20 mouse tissue samples) using the  
470 conventional human/manual-based approach. Notably, the time estimates for the exact same  
471 datasets varied from 1 minute (for 1 sample) to 18 minutes (for 20 mouse tissue samples) using the  
472 computer/MVP-based approach, thereby accelerating the analysis and figure/table generation  
473 process by ~22-fold on average (**Figure 4A**). Thus, MVP represents a QC software package in  
474 MS-based immunopeptidomics and provides unprecedented speed for fast visualization and quality  
475 assessment of immunopeptidomic data. Given its high-speed performance, MVP should find utility  
476 in QC analysis of large immunopeptidome sample cohorts.

477

**478 QC analysis of a large immunopeptidomic sample cohort**

**Running title: MhcVizPipe software**

479 To assess the performance of MVP to run 'batch' analysis, we selected a large immunopeptidomic  
480 dataset recently published in Shraibman et al. (46). In this study, Shraibman et al. report the analysis  
481 of the HLA immunopeptidome of 152 samples, including i) plasma-soluble HLA molecules (sHLA)  
482 of 142 plasma samples from both glioblastoma and non-cancerous patients, and ii) membranal  
483 HLA (mHLA) of 10 glioblastoma tumor tissues. This analysis covers 52 different HLA allotypes,  
484 more than 35,000 different HLA-associated peptides, and represents, to the best of our knowledge,  
485 the largest immunopeptidomic dataset ever generated from plasma samples (46). The 'batch  
486 analysis' of the 152 samples was run automatically using the command line interface and a simple  
487 Python script (<https://github.com/CaronLab/MhcVizPipe/wiki/Command-line-interface>). The run  
488 time was approximately 81 minutes (running in Ubuntu 20.04 on an 8-core desktop). The metrics  
489 files generated when running MVP from the command line (sample\_metrics.txt) were used to  
490 generate a QC plot showing three points/scores per sample (i.e. the total number of peptides  
491 detected, the LF and the BF scores) and illustrating the variability in these scores between all the  
492 samples (**Figure 4B**). In this dataset, the average LF score was 0.96 and the average BF score was  
493 0.80, highlighting the overall high quality of this large dataset. In addition, the generated QC plot  
494 was particularly useful to point out potentially problematic samples. For instance, among the lowest  
495 BF scores in this dataset was from the sample BCN\_13-Plasma\_A (**Figure 4C**). In fact,  
496 visualization of the data from this sample indicates that 151 peptides were predicted to be SB or  
497 WB whereas 376 peptides were predicted to be NB, hence, resulting in a relatively low BF score of  
498 0.29 (**Figure 4C**). Thus, MVP can run 'batch' analysis relatively rapidly to provide an assessment  
499 of the quality of individual samples within a large immunopeptidomic dataset. Together, MVP is  
500 a generic software tool, can be applied to any HLA immunopeptidomes, and given its high-speed

501 performance and scalability, will find utility to rapidly warn the user about problematic samples in  
502 large-scale clinical immunopeptidomic studies.

503

#### 504 **DISCUSSION**

505 The development of QC software is common practice for high quality research in the life sciences.  
506 Here, we present MVP, a GUI-based QC approach to rapidly and simultaneously assess the quality  
507 and MHC-specificity of small to large immunopeptidomic datasets generated by MS. Although  
508 MVP can be perceived as a relatively basic software solution by experts in the field, this is, to the  
509 best of our knowledge, one of the first reports focusing on the development of QC software in MS-  
510 based immunopeptidomics. Indeed, very little has been described in the literature with regard to  
511 the development of QC approaches, despite the fact that QC will become critically important in  
512 MS-based immunopeptidomics to comply with pharmaceutical regulations for real-life clinical  
513 applications (37).

514 For various reasons, the development of QC software tools has not been prioritized in  
515 immunopeptidomics over the last recent years. The main reason is most likely because most  
516 immunopeptidomic studies still have to deal with relatively small-scale datasets; and the quality  
517 and MHC-specificity of such datasets can still be manually assessed using the currently available  
518 peptide sequence clustering and MHC peptide binding algorithms. However, with the recent  
519 development of automated technologies for robust and high-throughput MHC immunoprecipitation  
520 protocols, the production of immunopeptidomic data on the scale of multiple gigabytes per  
521 instrument per day can be anticipated to accelerate the pace of capturing therapeutically relevant  
522 data (57). In this regard, at least two large-scale immunopeptidomics studies have recently been  
523 reported: 1) the profiling of immunopeptidomes from 152 clinical biospecimens, which were used

524 in this study (46), and 2), the mapping of the first draft of the human immunopeptidome from 227  
525 benign tissue samples, for a total of nearly 3,500 raw files generated by MS (47). With this  
526 perspective in mind, one can envision that other large-scale immunopeptidomic cohort studies will  
527 be conducted in the future, as recently discussed in the context of the human immunopeptidome  
528 project (14). Therefore, we believe that automated QC software solutions specialized in MS-based  
529 immunopeptidomics are inevitable. The creation of MVP represents a very first step in this  
530 direction and should stimulate the development of additional QC software solutions and products  
531 specialized in MS-based immunopeptidomics in the future.

532         The current version of MVP has obvious limitations. MVP was not designed to provide  
533 new biological insights. MVP is intended so far for situations where instrument operators or  
534 scientists in charge of downstream data analysis want a quick and consolidated view of the general  
535 quality, composition and MHC-specificity of one or multiple immunopeptidomic samples in  
536 parallel. Whether the user should continue using the evaluated data or not (e.g. with low quality  
537 data) will very much depend of the goal of the project, or the specific scientific question(s) asked.  
538 Therefore, we think that it is up to the user to judge what is best, if he/she should drop low quality  
539 samples or keep troubleshooting them to perform additional experiments and analyzes.

540         MVP is also subject to the accuracy of the available MHC peptide prediction algorithms  
541 and clustering tools. In this regard, new algorithms continue to be released for the prediction of  
542 HLA class I peptides (42, 43, 58) and HLA class II peptides (45, 59), and will be integrated in  
543 future versions of MVP in a yearly basis to improve the overall performance of the QC software.  
544 The predictors used in the future to upgrade MVP will also improve its accessibility. In the current  
545 version of MVP, NetMHCpan suite tools were used, although they remain poorly compatible with  
546 Windows. Therefore, installing MVP on Windows is possible, as explained in detailed in GitHub

**Running title: MhcVizPipe software**

547 (<https://github.com/CaronLab/MhcVizPipe/wiki/Windows-installation>) but is more challenging  
548 for a non-computer expert. Integration of Windows-compatible HLA peptide binding algorithms,  
549 such as MHCFlurry, will help to improve the accessibility of future versions of MVP.

550 Over the last recent years, considerable advances in HLA ligand purification protocols, MS  
551 instrumentations and computational methods have pushed forward the boundaries of MS-based  
552 immunopeptidomics. For instance, state-of-the-art data-independent acquisition (DIA) methods  
553 have been applied (39, 60, 61), specialized computational workflows have been created (62, 63),  
554 and new automated and semi-automated HLA ligand purification platforms have been developed  
555 to increase the throughput, speed, sensitivity and reproducibility of immunopeptidome analysis by  
556 MS (57, 64, 65). In the context of this progressive technological development, the current version  
557 of MVP will inevitably need to be further developed to reach its full QC capabilities and potential  
558 for the field. Currently, MVP is best suited for peptide datasets generated by data-dependent  
559 acquisition (DDA)-MS. However, a long-term goal in immunopeptidomics is to scale to large and  
560 quantitative clinical datasets generated by DIA-MS (39, 60, 61), and subject to stringent QC, as is  
561 performed in large-scale clinical proteomics studies (66–70). In this regard, generation of high-  
562 quality, quantitative immunopeptidomic data in large-scale studies is challenging because of issues  
563 induced by the samples themselves due to collection, storage and processing. To help overcome  
564 this challenge, MVP could be further developed to evaluate immunopeptidomic data in near real-  
565 time—as described in proteomics (27)—allowing for interventions as soon as deviations in data  
566 quality of specific immunopeptidomic samples are detected.

567 Beside QC software, we also believe that future community-driven immunopeptidomic  
568 studies that focus on the evaluation and establishment of QC samples/standards are needed. In fact,  
569 QC samples in immunopeptidomics remain poorly documented to date (37). Nevertheless, such

**Running title: MhcVizPipe software**

570 samples are critical in the long run to generate reproducible clinical datasets across multiple  
571 laboratories (13). If successful in doing so, robust and high-quality immunopeptidomic data would  
572 be consistently generated, stored and shared through specialized public repositories—e.g.  
573 SystemMHC Atlas (71, 72) and caAtlas (73)—thereby providing a global and sustainable mechanism  
574 to foster collaborations among computational scientists, biostatisticians, immunologists and  
575 clinical investigators to improve T-cell based immunotherapies (13).

576 In conclusion, MVP is a semi-automated, GUI-based QC software for rapid assessment of  
577 multiple immunopeptidomic datasets generated by MS. The current version of MVP parallelizes  
578 the use of well-established immunopeptidomic algorithms as well as new in-house software tools  
579 to assess data quality and MHC-specificity at unprecedented speed, and will therefore be of  
580 immediate utility for any expert and non-expert in the field. We envision that further development  
581 of this software tool will facilitate QC for upcoming large-scale immunopeptidomic cohort studies.

582

**583 DATA AVAILABILITY**

584 All of the data generated or analyzed during this study are included in this published article and/or  
585 the supplementary materials. Created datasets and code are publicly available. Immunopeptidomic  
586 data visualized in Supplemental Data S2 have been deposited to the ProteomeXchange Consortium  
587 (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset  
588 identifier PXD028633. All the codes are available at github:  
589 <https://github.com/CaronLab/MhcVizPipe>.

590

**591 ACKNOWLEDGMENTS**

592

**Running title: Mincvizipe software**

593 This work was supported by funding from the Fonds de recherche du Québec – Santé (FRQS), the  
594 Cole Foundation, CHU Sainte-Justine and the Charles-Bruneau Foundations, Canada Foundation  
595 for Innovation, the National Sciences and Engineering Research Council (NSERC) (#RGPIN-  
596 2020-05232) and the Canadian Institutes of Health Research (CIHR) (#174924). IRIC proteomics  
597 facility is a Genomics Technology platform funded in part by the Canadian Government through  
598 Genome Canada. KK is a recipient of IVADO's postdoctoral scholarship (#4879287150).  
599  
600

## 601 REFERENCES

602

603 1. Wan, X., Vomund, A. N., Peterson, O. J., Chervonsky, A. V., Lichti, C. F., and Unanue, E. R.  
 604 (2020) The MHC-II peptidome of pancreatic islets identifies key features of autoimmune peptides.  
 605 *Nat Immunol*, 21, 455-463

606 2. Gonzalez-Duque, S., Azoury, M. E., Colli, M. L., Afonso, G., Turatsinze, J.-V., Nigi, L., Lalanne,  
 607 A. I., Sebastiani, G., Carré, A., Pinto, S., Culina, S., Corcos, N., Bugliani, M., Marchetti, P.,  
 608 Armanet, M., Diedisheim, M., Kyewski, B., Steinmetz, L. M., Buus, S., You, S., Dubois-Laforgue,  
 609 D., Larger, E., Beressi, J.-P., Bruno, G., Dotta, F., Scharfmann, R., Eizirik, D. L., Verdier, Y., Vinh,  
 610 J., and Mallone, R. (2018) Conventional and Neo-Antigenic Peptides Presented by  $\beta$ 2-Microglobulin  
 611 Are Targeted by Circulating Naïve CD8<sup>+</sup> T Cells in Type 1 Diabetic and Healthy Donors. *Cell*  
 612 *Metab* 28, 946-960.e6

613 3. Gubin, M. M., Zhang, X., Schuster, H., Caron, E., Ward, J. P., Noguchi, T., Ivanova, Y., Hundal,  
 614 J., Arthur, C. D., Krebber, W.-J., Mulder, G. E., Toebes, M., Vesely, M. D., Lam, S. S. K., Korman,  
 615 A. J., Allison, J. P., Freeman, G. J., Sharpe, A. H., Pearce, E. L., Schumacher, T. N., Aebbersold,  
 616 R., Rammensee, H.-G., Melief, C. J. M., Mardis, E. R., Gillanders, W. E., Artyomov, M. N., and  
 617 Schreiber, R. D. (2014) Checkpoint blockade cancer immunotherapy targets tumour-specific  
 618 mutant antigens. *Nature* 515, 577-581

619 4. Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub,  
 620 M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., Busch,  
 621 D. H., Peschel, C., Rad, R., Cox, J., Mann, M., and Krackhardt, A. M. (2016) Direct identification  
 622 of clinically relevant neoepitopes presented on native human melanoma tissue by mass  
 623 spectrometry. *Nat Commun* 7, 13404

624 5. Yadav, M., Jhunjhunwala, S., Phung, Q. T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C.,  
 625 Cheung, T. K., Fritsche, J., Weinschenk, T., Modrusan, Z., Mellman, I., Lill, J. R., and Delamarre,  
 626 L. (2015) Predicting immunogenic tumour mutations by combining mass spectrometry and exome  
 627 sequencing. *Nature* 515, 572-576

628 6. Paes, W., Leonov, G., Partridge, T., Chikata, T., Murakoshi, H., Frangou, A., Brackenridge, S.,  
 629 Nicastrì, A., Smith, A. G., Learn, G. H., Li, Y., Parker, R., Oka, S., Pellegrino, P., Williams, I.,  
 630 Haynes, B. F., McMichael, A. J., Shaw, G. M., Hahn, B. H., Takiguchi, M., Ternette, N., and  
 631 Borrow, P. (2019) Contribution of proteasome-catalyzed peptide cis-splicing to viral targeting by  
 632 CD8<sup>+</sup> T cells in HIV-1 infection. *Proc Natl Acad Sci USA*. 116, 24748-24759

633 7. Croft, N. P., Smith, S. A., Pickering, J., Sidney, J., Peters, B., Faridi, P., Witney, M. J., Sebastian,  
 634 P., Flesch, I. E. A., Heading, S. L., Sette, A., Gruta, N. L. L., Purcell, A. W., and Tschärke, D. C.  
 635 (2019) Most viral peptides displayed by class I MHC on infected cells are immunogenic. *Proc Natl*  
 636 *Acad Sci USA* 116, 3112-3117

637 8. Wu, T., Guan, J., Handel, A., Tschärke, D. C., Sidney, J., Sette, A., Wakim, L. M., Sng, X. Y.  
 638 X., Thomas, P. G., Croft, N. P., Purcell, A. W., and Gruta, N. L. L. (2019) Quantification of epitope

866 **FIGURE AND TABLE LEGENDS**

867

868 **Figure 1. Overview and installation of the MhcVizPipe (MVP) software.** (A) Illustration  
869 showing the experimental workflow for the isolation and identification of MHC-associated  
870 peptides by LC-MS/MS (i), the software tools integrated into the MVP pipeline for rapid data  
871 processing (ii), and the HTML report generated to support (or not) downstream immunopeptidomic  
872 data interpretation through quality control (QC) and quality assurance (QA) (iii). The MVP pipeline  
873 was created with the Python language and parallelizes in-house (UpsetPlotly and Plotly-Logo) and  
874 established (NetMHCpan, NetMHCIIpan and GibbsCluster) algorithms. (B) Installation steps. The  
875 user can download and install the MVP software from <http://github.com/CaronLab/MhcVizPipe>  
876 on Linux, Mac and Windows 10 (using the Windows Subsystem for Linux). The installation  
877 requires the download of the third-party software tools (i) and the installation of MVP (ii).

878

879 **Figure 2. Overview of the MVP graphical user interface (GUI).** To run an analysis, the user can  
880 copy-paste a list of peptides or upload more than one file at a time. Samples can be labeled with  
881 detailed information. The user click the "GO!" button to start an analysis and a loading screen  
882 appears while the analysis is running, followed by a pop-up window with a link to the HTML report.  
883 More details about the GUI are available at <https://github.com/CaronLab/MhcVizPipe/wiki/Usage>.  
884 A command line interface (CLI) is also available for 'batch' analyses at  
885 <https://github.com/CaronLab/MhcVizPipe/wiki/Command-line-interface>.

886

887 **Figure 3. Visualization of the QC scores (LF & BF) calculated in Table 1.** (A,B) Histograms  
888 and heatmaps illustrating high quality (A) and middle to low quality (B) immunopeptidomic data

**Running title: Mincvizipe software**

889 generated by MS from various tumor biopsies (see Table 1). Histograms showing the distribution  
890 of peptides according to their length (right panel) and predicted MHC binding affinity (left panel).  
891 SB: Strong Binder, WB: Weak Binder, NB: Non-Binder. Heatmaps are automatically generated by  
892 MVP and are provided in the HTML report. For each HLA allele, NetMHCpan provides a %Rank  
893 score for individual peptides, which are color coded on the heatmap in red, blue and yellow for SB,  
894 WB and NB, respectively, with a linear gradient between the colors.

895

896 **Figure 4. Speed performance of the MVP software for small- to large-scale QC analyses.** (A)  
897 Pie charts showing the time estimates to process and analyze the selected datasets manually  
898 (human-based in green) or using MVP (computer-based in red). Datasets tested were: (from left to  
899 right) 1 QC sample (JY), 1 PBMC sample, 15 PBMC samples, 5 mouse tissue samples and 20  
900 mouse tissue samples. (B) Large-scale QC analysis of the HLA peptidome of 152 samples. Graph  
901 showing the total number of peptides and the two scores calculated based on peptide lengths (LF)  
902 and the number of peptides predicted to bind HLAs (BF) (y-axis). Each sample has three points,  
903 one for each of the values (shown in the legend). The points for each sample are connected by a  
904 vertical line. (C) Zoom in on the sample 'BCN\_13-Plasma\_A'. Histogram and heat map showing  
905 the length distribution and predicted HLA binding affinity (%Rank) of the detected peptides in this  
906 sample. LF and BF scores are indicated.

907

908 **Table 1. Length Fraction (LF) and Binding Fraction (BF) scores calculated from**  
909 **immuno-peptidomic data generated from eight different tumor biopsies.** Immuno-peptidomic  
910 data were obtained from Rijensky et al. (2020). Eight different biopsies and cancer types were  
911 analyzed. The sample weight, the total number of peptides identified, the fraction of peptides

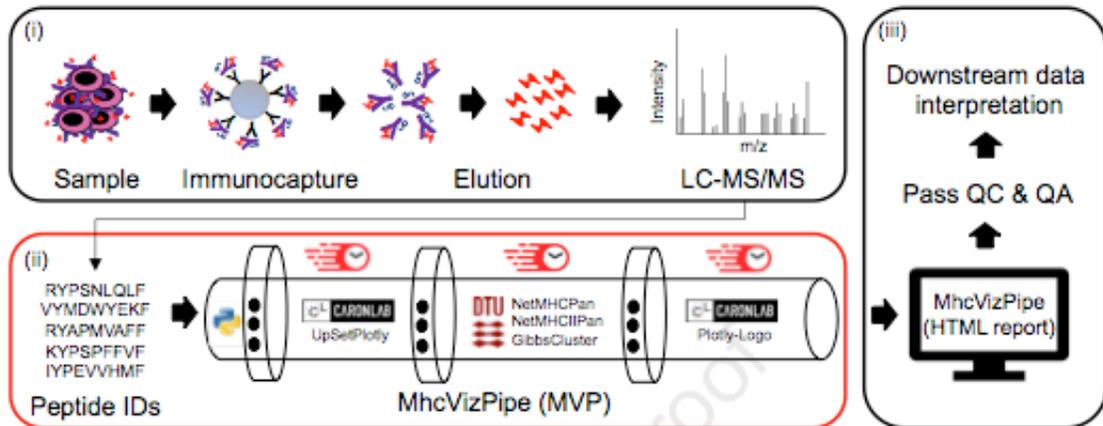
**Running title: MincVizipe software**

912 between 8 and 12 mers and the LF and BF scores are indicated. LF and BF scores are color-coded  
913 to illustrate high quality (green), middle quality (brown-red), and low quality (red)  
914 immunopeptidomic data.

Journal Pre-proof

Figure 1

A



B

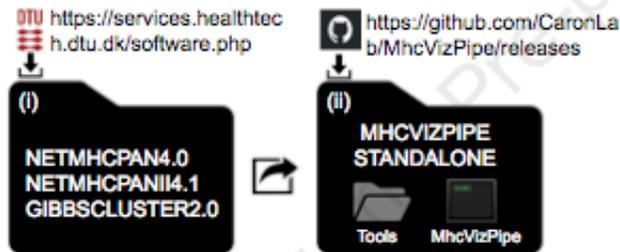


Figure 2





A quick and user-friendly visualization tool for mass spectrometry data of MHC class I and II peptides.  
[Click here for help and resources](#) CHECK FOR UPDATES SETTINGS

---

To load data, use the light blue area below or copy-and-paste into the text box below (you may load any number of samples):

Drag and drop one or more files in this area, or [CLICK TO SELECT FILE\(S\)](#) (.txt, .csv, .tsv)

Note: you cannot sequentially load files with identical contents (i.e. duplicates).

Paste a peptide list or select a file using the above interface

**Sample information:**

Sample name

Sample description (optional)

LOAD DATA ↩

**Loaded data:**

Sample name	Description (optional)	Alleles (required)
- biopsy_F2		HLA-A2402, HLA-B5801, HLA-C6701
- biopsy_F4		HLA-A2381, HLA-A8281, HLA-B4483, HLA-B3881, HLA-C8801, HLA-C1293

**Allele search (click arrow to add to selected cells):**  
Select...

---

**MHC class:**

Class I
x
v

**General information:**

Experiment description (optional)

Submitter name (optional)

**Experimental information (optional):**

Species:

# of cells:

Lysis buffer:

Type of beads:

Antibody:

Incubation time:

MHC ligand complex elution buffer:

Peptide elution buffer:

Type of MS/MS:

Peptide identification software:

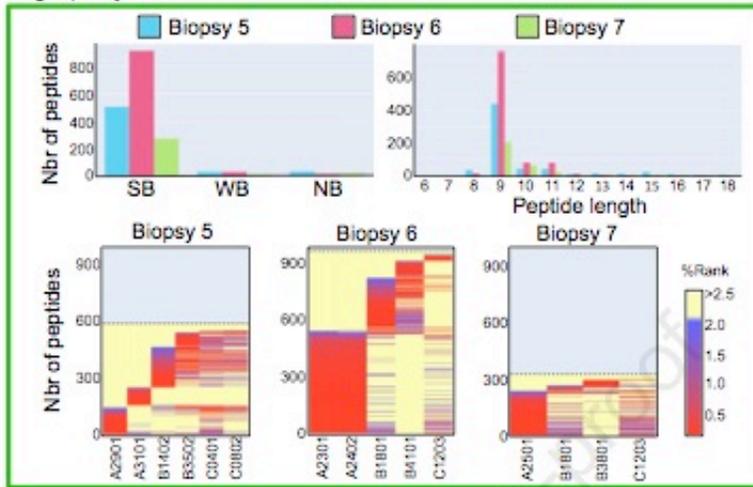
Peptide FDR:

(Enter any further information using the same format)

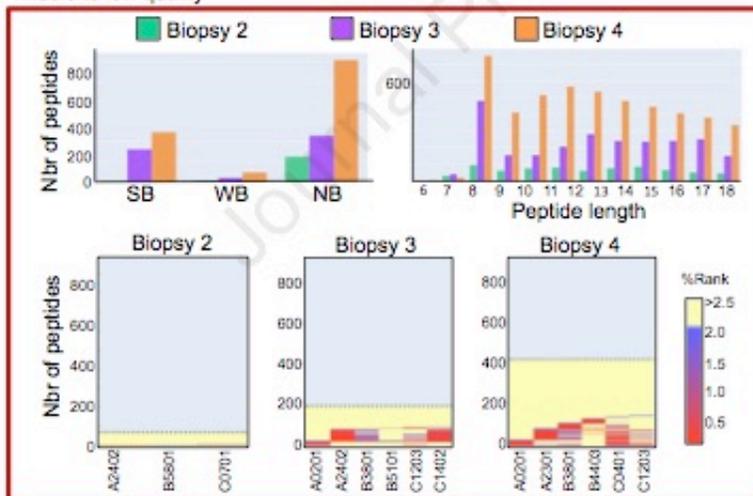
GO!

**Figure 2**

**A** High quality

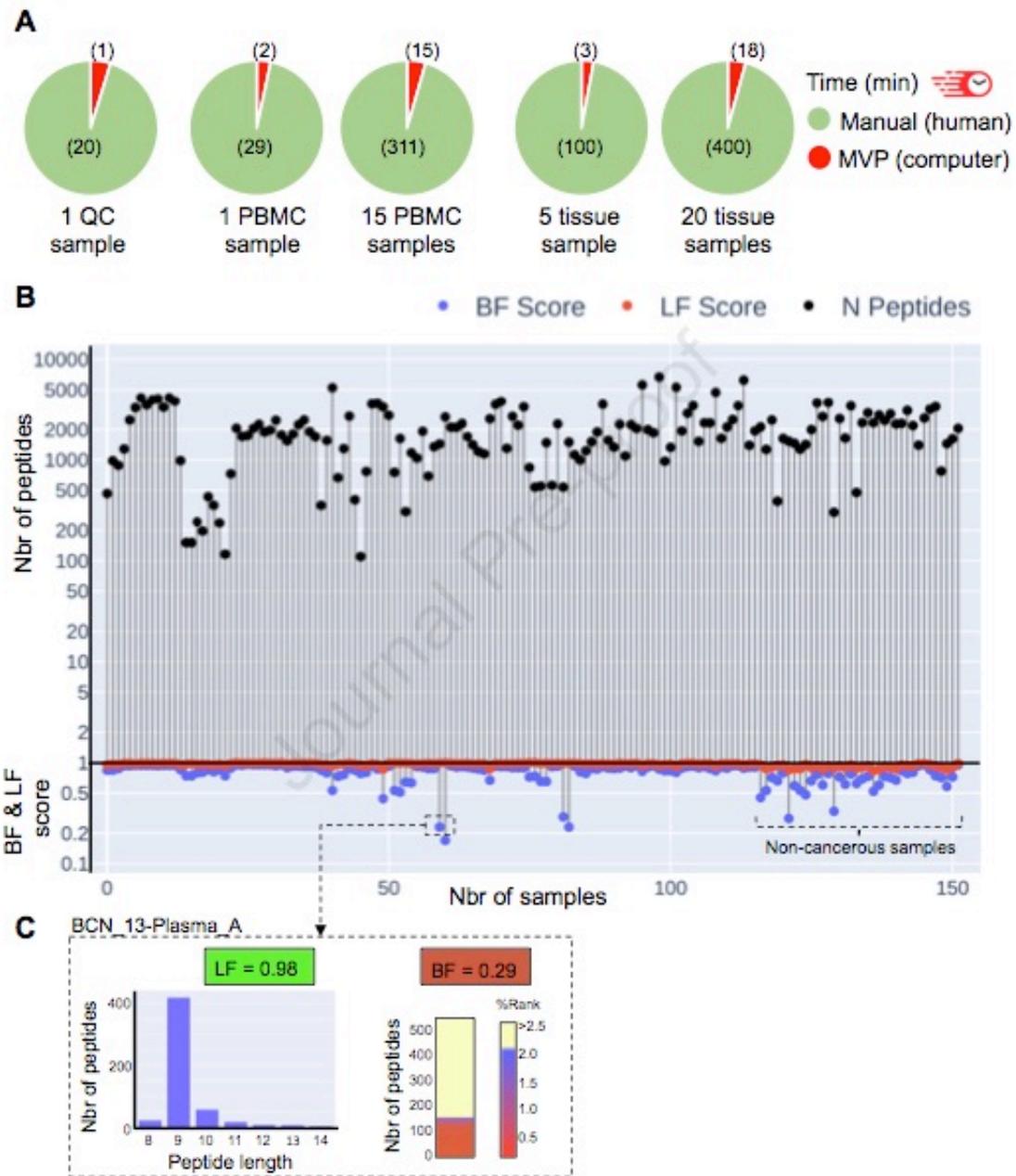


**B** Middle to low quality



**Figure 4**

Journal Pre-proof



## Annexe 5 - Immunopeptidome des cellules MRC5 naïves et infectés par Coronavirus OC-43

**Tableau 1** - Immunopeptidome des cellules MRC5 naïves (protéome humain)

Peptide	Longueur	Source	PTM
KLLPALLLQHV	11	P14210	
RPDGEKKAYVRL	12	P62750	
YFYDPDVGNFHY	12	O15379	
LM(+15.99)LDPHVQGGFY	12	Q12952	Oxidation (M)
GPPSHSLRLPL	12	Q96AY3	
VFIVNSEDYM(+15.99)Y	11	O75182	Oxidation (M)
EEHLESHKKY	10	Q9Y2D5	
FYEVFKVLY	9	Q00325	
FYAFQDDRYLY	11	O75116	
HYFSDPFLKY	10	Q15165	
AEIGEGAYGKVF	12	Q00534	
KLLEGEESRISL	12	P08670	
YYYDGDVGNYYY	12	Q13547	
SLFPHPHIHEY	11	Q5W0V3	
APRPASGPIRP	11	P26373	
RLYHEADKTYM(+15.99)	11	O75663	Oxidation (M)
TIDDPTKQTKF	11	Q96RF0	
EEVHDLERKY	10	P55209:Q99733	
ALFPHPHIHEY	11	Q86V87	
SM(+15.99)YGVDLHHA	10	O43491:Q9Y2J2:Q9H4G0	Oxidation (M)
AEIRIIPAKKY	11	Q15392	
GM(+15.99)YDQAVKY	9	Q96AY4	Oxidation (M)
VM(+15.99)M(+15.99)DPNSTQRY	11	Q9HAV4	Oxidation (M)
AEDKENYKKFY	11	P08238	
FFGETSHNY	9	P43243	
AEDKENYKKF	10	P08238	
AALEEADKM(+15.99)FL	11	Q9Y6B2	Oxidation (M)
GPNERIGM(+15.99)V	9	P52292	Oxidation (M)
YFSDPFLKY	9	Q15165	
VPDSSGPERIL	11	P61978	
EEQQALQKKL	10	P07951	
LPYNHQHEY	9	O95864	

EEIAIIPSKKL	11	P08708	
REIQDEELKKF	11	O43299	
HFIAEHYM(+15.99)F	9	O15121	Oxidation (M)
RPKETDHKSL	10	Q9BVI0	
RPRAEVRFL	10	Q13438	
YFEYIEENKY	10	Q00839	
NPAENFRVL	9	P13010	
EALQDVEDENQ	11	P62258	
FLASESLIKQI	11	P62906	
YM(+15.99)ADLESALHY	11	O00391	Oxidation (M)
TEANVVRKF	9	O00391	
KYFDEHYEY	9	P33552	
FRQPKVVRVY	9	Q13219	
NEM(+15.99)QPNGKVF	10	O60488	Oxidation (M)
FFPDKPITQY	10	P39656	
FIHTDPSRFM(+15.99)Y	11	Q8IVL0	Oxidation (M)
HVLIDVGTGY	11	Q99471	
KEGLDIYKKF	10	Q13492	
NYLHDFLKY	9	Q9UBU8	
YYYDGDIGNYYY	12	Q92769	
M(+15.99)M(+15.99)NGGHYTY	9	Q99497	Oxidation (M)
YWGEIISQQY	10	P14735	
TEEEKNFKAF	10	P26373	
SLLQATDFM(+15.99)SL	11	P04216	Oxidation (M)
VM(+15.99)YNSSDHDVY	11	Q9H0E9	Oxidation (M)
KPAPAKPVAA	10	P12111	
GGPGTASRPSS	12	P08670	
RPWLEGRHTL	10	Q9UI10	
YAFQDDRYLY	10	O75116	
M(+15.99)M(+15.99)DPNSTQRY	10	Q9HAV4	Oxidation (M)
AEHLGDFHRTY	11	Q92530	
NNGQRLTY	9	Q8WX93	
SEGVRPDGF	9	P12111	
SEM(+15.99)NTDKQYF	10	Q96D46	Oxidation (M)
YFLGIKSLVY	10	O15121	
RPRPGTGLGRVM(+15.99)	12	Q9BVK6	Oxidation (M)
GEDVETSKKW	10	O60869	
FYLSTPENYHY	11	Q9HD67	
NEDNGIIFKAF	10	P36578	
NPNQKNVAL	10	Q15717	
RPVQKLIRL	9	Q5W0V3	

RPVQKLLRL	9	Q86V87	
EELDFQKNIY	10	P02545	
GYYGKGLQY	9	Q4LDE5	
RPDGEKKAYV	10	P62750	
M(+15.99)VTEIRLKY	9	Q15393	Oxidation (M)
TFM(+15.99)DHVLRV	9	P53396	Oxidation (M)
AEKAYHEQL	9	Q9BQE3:Q71U36:P68363:PODPH7:Q6PEY2:PODPH8	
NWNEVVTQQY	10	O43847	
SPRLDHTRAV	10	Q00610	
RVDPNGSRVLL	11	Q16531	
SASVQRADTSL	11	O76080	
AASKERSGVSL	11	P16402	
HVIDVKFLY	9	Q16531	
KLLDISELDM(+15.99)V	11	Q9H3P2	Oxidation (M)
SEAEIIRKY	10	P09874	
SEGDIQKGY	10	Q9P265:Q9Y2E4:Q14689	
KVLDFEHFL	9	P60660	
AELDRQIKSF	10	Q13813	
RLNEGDLHGM(+15.99)	11	O75762	Oxidation (M)
YYRYPTGESY	10	Q16875	
HPRIITEGF	9	P40227	
SPRLPVGGF	9	Q14669	
AEDENGKIVGY	11	P41227	
APRPSTGPHKL	11	P62701	
LFETPVEAQY	10	Q08431	
ALWDIETGQQTV	12	P62879	
AEKELHEKF	9	Q9NVA2	
EESDLRQY	9	Q12789	
GNYEKVFEY	9	Q92624	
RPHETGGYM(+15.99)L	10	P02751	Oxidation (M)
YYINKISSTLY	11	Q8N6L1	
EAM(+15.99)EDGEIDGNK	12	P19338	Oxidation (M)
VAYDSKQYY	9	Q96KR1	
SPNAEIHIL	9	Q14807	
LM(+15.99)LPEKLRPLY	11	O95563	Oxidation (M)
LPRPPPEM(+15.99)	9	Q00325	Oxidation (M)
QPM(+15.99)AAPEGKRSL	12	Q9Y6B2	Oxidation (M)
HVLDVIHEV	9	O75962	
FVYHNDIFRY	10	Q13283	
GPNERIGM(+15.99)VV	10	P52292	Oxidation (M)
APTIVGKSSL	10	P39656	

EEIAFLKKL	9	P08670	
AEGIHTGQF	9	P62917	
SPQSGNFRQL	10	Q9H074	
SFAARSFYY	9	P49327	
SPAHRPALL	9	P78539	
SPAKNPSSL	9	Q9GZR7	
YPDGSKPVKV	10	Q01995	
TSFDKFKYK	9	P45985	
VLFEHAVGY	9	O00567	
YYIFIPSKF	9	Q9H0S4	
HLKNSQEKY	9	P53634	
FFKPHWDEKF	10	P50454	
HLDLPSNNNL	10	P18583	
GIKNFFTDVY	10	P0DI82:P0DI81	
GYLHDFLKY	9	Q15014	
FLNDSYLKY	9	Q8WVM7:Q8N3U4	
TLFSEPKFAQV	11	Q9H425	
EETETVTKSY	10	Q05682	
GEVLISRVY	9	Q96CW1	
VAISEELVQKY	11	Q9NQC3	
GLM(+15.99)GPRGPPGA	11	P08123	Oxidation (M)
DPM(+15.99)KARVVL	9	P49458	Oxidation (M)
AM(+15.99)SSKFFLV	9	P41221	Oxidation (M)
IQDNHDGTYTV	11	P21333	
DEIERKFDKW	10	Q8WWY3	
ADIDTRSEFY	10	P61160	
SFAEIIIM(+15.99)GY	9	P40763	Oxidation (M)
REHDANKINY	10	O76074	
SM(+15.99)TLAIHEI	9	O15121	Oxidation (M)
RVQKIRPSI	9	Q9H3M7	
AETGQVQKI	9	Q00610	
FFGM(+15.99)HVQEY	9	Q09472	Oxidation (M)
AYSSYVHQY	9	Q92624	
FFHDILHTKY	10	Q92508	
SM(+15.99)M(+15.99)DPNHFL	9	Q8IWW8	Oxidation (M)
RPDDKRILEF	10	O60502	
SPDPKASQHF	10	P12111	
AEKLITQTF	9	Q9Y266	
TPKM(+15.99)PGQSV	9	Q8WYP5	Oxidation (M)
EEAQFETKKLY	11	O60613	
KPTLPIKTF	9	O00254	

GPDSTRHTTL	10	Q14582	
KM(+15.99)YEEFLSKV	10	P10644:P31321	Oxidation (M)
KPHNPGFGM(+15.99)	9	Q12906	Oxidation (M)
SESLLRGIYAY	11	P60842	
SPRLPVGGFRSL	12	Q14669	
YFEEIKQFY	9	Q7KZ85	
VLLDKIKNLQV	11	P12111	
KEM(+15.99)EQFVKKY	10	Q9UGI8	Oxidation (M)
LARPPPASA	9	Q9HCN3	
FRPLYTHEY	9	Q99442	
TM(+15.99)VGKTFTY	9	Q16850	Oxidation (M)
KLYTKHGTLKY	11	P69849:Q15155:Q5JPE7	
NESLFGKKY	9	Q14192	
AIDQLHLEY	9	Q43707	
GPSKVRVGVV	10	P12111	
YFLRPM(+15.99)LQY	9	O00410	Oxidation (M)
APSLFHLNTL	10	Q96QB1	
SEPDFVAKFY	10	Q9NWQ9	
FM(+15.99)FDEKLVTV	10	O00743	Oxidation (M)
APSQGRPGLSL	11	P12109	
APRAPPGPAGL	11	Q9NRL3	
GTSSYTAQEY	10	Q92841	
YFHHQDGYRY	10	Q92802	
APRQPGLM(+15.99)AQM(+15.99)	11	Q9Y6H1:Q5T1J5	Oxidation (M)
RPHYGKLKEI	10	Q13200	
TPNKETAGL	9	Q05682	
REDVQKKTFF	9	P11532	
EELRAHVRKY	10	Q9Y2A7	
GLWEDGRSTLL	11	Q96BA8	
SPTVRVSVV	9	P12111	
NEDPAKRPF	10	Q13418	
TAEDPSKSYV	10	O14929	
APRPDRLVNRL	11	O75534	
RVNDVPEEFM(+15.99)Y	11	O95487	Oxidation (M)
QLYDKGLVY	9	P41252	
VM(+15.99)DSKIVQV	9	O15131:O60684:P52294	Oxidation (M)
M(+15.99)PVGPDAILRY	11	P46379	Oxidation (M)
VFNEKGWNY	9	P43490	
KLIDIFQKY	9	Q92484	
SPQPAQQKL	9	Q14157	
EAEIIRKY	9	P09874	

GVM(+15.99)DYLLKY	9	O95365	Oxidation (M)
LEDGRTLSDY	10	P62979	
GPREFRQL	9	P57086	
VLLEQEKTFFT	11	Q9NS68	
APTGSCKTI	9	O75643	
APTGSCKTL	9		
PRQPPM(+15.99)SL	8	Q13045	Oxidation (M)
NEDEDSPNKLY	11	P62899	
YYNGKIHAY	9	O75643	
Q(+.98)ELQEINRVY	10	P07355	Deamidation (NQ)
EEALHASHGF	10	Q9P0J0	
SEVILHHEY	9	O75643	
AADRPVSL	9	A1A456	
TPSLVKSTSQL	11	Q06055	
TRVKAVRIL	9	Q9H3M7	
SEKAGIIQDTW	11	Q9P2E9	
M(+15.99)TNPVHLY	9	Q15006	Oxidation (M)
EEFELLGKAY	10	Q99613:B5ME19	
APEEHPTLL	9	P68133:P68032:P63267:P62736	
ALTGHLEEV	9	P04083	
IM(+15.99)ATPEQVGKM(+15.99)	11	P12956	Oxidation (M)
YLLESAKM(+15.99)FNY	11	Q96G23	Oxidation (M)
YPFKPPKVAF	10	P62837:P61077:Q9Y2X8	
SPDIRNTTV	9	Q96PD2	
SFATLFPKY	9	Q13601	
ALLDTTEKYLY	11	Q8N587	
SEAKAFHDY	9	Q9BWW4:P81877	
GM(+15.99)YGKIAVM(+15.99)EL	11	Q16531	Oxidation (M)
ASYEVKELY	9	O43252	
SRFLNKQPY	9	P10619	
APPEYHRKAV	10	Q9UBU8	
AEAEAVREVV	10	Q9BZJ0	
VLLKARLVPA	10	Q6P995	
APNNGQNKSL	10	Q9UKL0	
FFPSIYDVKY	10	Q9UFF9	
AEIVEGENHTY	11	P21333	
RQADKVVRL	9	P08651:Q12857:Q14938:O00712	
YTM(+15.99)DDAWKY	9	Q8WW11	Oxidation (M)
TPSEHPVL	9	Q9Y5A9	
KLFIGGLNV	9	Q13151	
EEEAKAKGTF	10	Q9UHQ7	

AEFKEAFQL	9	P60660	
EESQEIQKKLY	11	O75116	
YLLPHILVY	9	Q13535	
KIAPNTPQL	9	P69849:Q15155:Q5JPE7	
AEEKQTKSF	9	Q00341	
SPRSPSGHPHV	11	Q8WX93	
VYKVPPFTF	9	Q6NZI2	
TM(+15.99)LARLASA	9	Q6UVK1	Oxidation (M)
QETSFTKEAY	10	P13693	
AEKAVTKEEF	10	P08865	
KEIDKNDHLY	10	Q9UNF1	
KIFEM(+15.99)GPVFTL	11	P00403	Oxidation (M)
SPLNTGKAL	9	P12111	
NTYPIKLFY	9	P61619	
GVDVTGPHLY	10	Q99436	
FSDDPKVEF	9	P12111	
GEVDYKKKY	9	Q8NBZ0	
SVSDQFYRY	9	P55010	
FIASKGVKLV	10	O43707:P12814	
TEHKSPDGRTY	11	O75400	
SNFRSNFGY	9	P25788	
NM(+15.99)VDIIHSV	9	P05166	Oxidation (M)
IDEPLEGSEDR	11	P61978	
HLHGKTFDY	9	Q08945	
KPAPAKPVA	9	P12111	
YVANEFLKY	9	Q14314	
KEIDKEEHLY	10	Q96JG8	
KPYVDGKGREL	11	Q13813	
SPSSVTGNAL	10	Q9NYB0	
NYSHLVSVDY	10	Q6ZMW2:Q9NYW8:Q13401:Q5JUW0:P17030: P17032	
FTYNGRTFY	9	P02751	
TTDGPTKEF	9	Q99715	
SEFIDSQRVW	10	Q6P2Q9	
QPRGGDYRGF	10	Q9Y5G6	
FM(+15.99)YDRPLRL	9	P12111	Oxidation (M)
GPGLSRTGVEL	11	P21333	
TPDTGRILSKL	11	A2A3N6	
FFPTQGHDY	9	Q8N3C0	
HPKKPEHTL	9	P32456:Q9H0R5	
SYLDKEDTFFY	11	Q9BTC8	
STVSRFFLY	9	Q969X6	

EVYAGSHQY	9	Q9H3P7	
APDQPLKAL	9	Q9BUX1	
KLHDVELHQV	10	O14775	
KLLTPITTL	9	P12111	
SEFTGVWKY	9	O75340	
HVDLDSGLRY	10	Q8WYQ9	
YALDYAQRV	9	P12111	
NVLEHFHKY	9	Q5VIR6	
SAYEALEKY	9	Q15785	
NSDTHSSTF	9	P04406	
SEINKPNFY	9	Q5QJE6	
APEEHPVLL	9	P60709:P63261	
SPVVRVAV	8	P13639	
KADTVSKTEL	10	Q9P0V9	
QEFQVPHQY	9	O60502	
LPRGSIPRSL	10	Q14566	
GEDEEENNLEVR	12	P04843	
SEYARPHVM(+15.99)	9	P54259:Q9P2R6	Oxidation (M)
NM(+15.99)YGKVVTV	9	O00267	Oxidation (M)
ALSDHHIYL	9	P04075	
EESQPAKKTY	10	O75400	
AEPYGGKIKNY	10	P43243	
EEPTVIKKY	9	Q9Y5X3	
SYHQQVQTY	9	Q12946	
DALREPGKAL	10	Q8IU81	
RPREAVERF	9	Q13438	
GVYGDVQRV	9	P26599	
RPYAAGGKKL	10	Q9H7U1	
STFFPFHEY	9	O60306	
VM(+15.99)WSPTHPALF	11	Q13409	Oxidation (M)
HYPPVPFGY	10	Q14201	
EGNPDTHSW	9	Q14157	
ILM(+15.99)EHIHKL	9	P84098	Oxidation (M)
YVIDPIKGLKL	11	Q15149	
RPKYPGIKF	9	P45984	
RPHETGGYM(+15.99)	9	P02751	Oxidation (M)
EEEKNFKAF	9	P26373	
LEGNPDTHSW	10	Q14157	
YM(+15.99)IAHITGL	9	P04818	Oxidation (M)
YLDIKGLLDV	10	P63208	
GSHSM(+15.99)RYFY	9	P01889	Oxidation (M)

EEM(+15.99)ARHLREY	10	P08670	Oxidation (M)
RTAHVILRY	9	Q9UKV5	
TENDIRVM(+15.99)F	9	Q92879	Oxidation (M)
AELTPHQTF	9	P04844	
RLGEAEGLM(+15.99)KV	11	Q9HBR0	Oxidation (M)
YKVEYPIM(+15.99)Y	9	O14880	Oxidation (M)
SEDEIKKAY	9	Q99615	
AEELERQGY	9	Q7KZ85	
APAARFNNSQF	11	Q99967	
EEINRQVTY	9	Q14517	
SSFKNYRY	9	Q9BTE7	
SPHFQHAEV	9	Q9Y2J4	
HEIEALKSF	9	O43847	
SNSQLM(+15.99)AVQY	10	P18850	Oxidation (M)
NLLEKDYFGL	10	O43491:Q9Y2J2:Q9H4G0	
NLLEKDYFGI	10		
TVLDVHKKY	9	Q13616	
RVPPPQSV	9	P29353	
NM(+15.99)VAKVDEV	9	P62906	Oxidation (M)
NFSHGTHEY	9	P14618	
KLYEKLEF	9	Q9BZJ0	
IAM(+15.99)EFNHLY	9	P23219	Oxidation (M)
GNYSGAAEYLY	11	P60228	
AEELGHTETY	10	Q9Y2G9	
EANVVRKF	8	O00391	
NAFEHGGEFTY	11	Q16829	
TVADKIHSV	9	O95147	
VGDPHTVTV	9	Q9H0E9	
YEVEIDGKTY	10	P50579	
GTVEKAYKY	9	Q8WXB4	
KEFDHPNVM(+15.99)	9	P30530	Oxidation (M)
AFFAERLYY	9	P20073	
ALSGHLETV	9	P07355	
VM(+15.99)APRTLLL	9	P04439	Oxidation (M)
SYHDPIM(+15.99)GNKY	11	O94915	Oxidation (M)
KLVSVAlM(+15.99)Y	9	P35318	Oxidation (M)
EEAKAKGTF	9	Q9UHQ7	
NPASKVIAL	9	Q00610	
SPNKLYTL	8	P62899	
APAGRKVGL	9	P62917	
NLM(+15.99)EQPIKV	9	P14923	Oxidation (M)

NYVEFTRIL	9	P24844	
SIGDIFLKY	9	Q9H8V3	
ALM(+15.99)FTGHLL	9	Q8IZU8	Oxidation (M)
KPGTFPKAL	9	Q96QB1:Q9Y3M8	
YPFKPPKVTF	10	Q969T4	
APAQRAQSL	9	Q63ZY3	
APRTPGRHGL	10	Q96M96	
RELPGHTGY	9	P62879	
AEFKEAFQLF	10	P60660	
EGLELPEDEEEK	12	P08238	
TLVYHVVG	9	Q15393	
SEFFRSGKY	9	P06733	
RPGFQQTSH	9	Q13443	
FPNIPGKSL	9	O15121	
LPAPPTQNM(+15.99)	9	Q15532:O75177	Oxidation (M)
HLGVKVFVS	9	P12109	
RLTPKLM(+15.99)EV	9	O15372	Oxidation (M)
VTDSNRSIATL	11	Q9Y678	
HLYIQM(+15.99)EY	9	Q9P2K8	Oxidation (M)
KLYPQLPAEI	10	Q9HCE0	
SPFSYKNVL	9	P05556	
RINHERNEL	9	O14777	
SGGNYRDNYDN	11	P98179	
YEQDQKYSF	9	P02751	
HVSTHQHNY	9	P01106	
AEESDLRQY	10	Q12789	
IPNEIIHAL	9	P52272	
NFYDETYDY	9	P61978	
RPDRGRLGF	9	Q01995	
QIVDVFM(+15.99)EY	9	Q00610	Oxidation (M)
KLLPGDIHQI	10	Q14185	
STYPRPHEY	9	Q9UBR2	
YQDPHSTAV	9	P00533	
SEEQRFNNF	9	P13010	
FFTSVSRPGR	10	P04439	
GEASRLAHY	9	Q99879:Q93079:Q99877:O60814:P62807:P58876:Q5QNW6	
GM(+15.99)YVFLHAV	9	Q9P0S3	Oxidation (M)
GVM(+15.99)DRGYSY	9	P28074	Oxidation (M)
TENDIYNFF	9	P31943:P52597	
VM(+15.99)APRTVLL	9	P01889	Oxidation (M)
EEM(+15.99)DLFTKY	9	Q969Q6	Oxidation (M)

EEQAVRPKY	9	P14635	
NEVVAGIKEY	10	Q9UBU8:Q15014	
NLLEKQLEY	9	Q86XR8	
RPDSAHKM(+15.99)L	9	Q8WX93	Oxidation (M)
SPANPAHIL	9	Q9H967	
RELEKDM(+15.99)HYY	10	Q7L523:Q5VZM2	Oxidation (M)
APSRNGM(+15.99)VL	9	P26373	Oxidation (M)
NPEKRYNVL	9	Q1KMD3	
SFVDPLVTNY	10	Q9H3H1	
NEM(+15.99)KKYIAY	9	P12111	Oxidation (M)
YPKRPLLGL	9	Q9GZR7	
RPKVPDQSV	9	Q8NCT1	
FNSHVHVEY	9	O43150	
GGYGGSRDYY	10	Q14011	
TYAGQFNQDY	10	P50995	
VVFEGNHYFY	10	P04843	
RPRSALPRLLLP	12	Q14332	
KLLPYIVGV	9	P12111	
GLNDFIQKI	9	O00141	
AIITTHLQY	9	Q96B45	
AEDEGVSQRKF	11	O00299	
DESGEGHKW	10	Q96K76	
REAVQKKTF	9	Q01082	
KPKDPLKISL	10	Q969Q6	
TVADHIQKV	9	Q9Y678	
GM(+15.99)YIFLHTV	9	Q8N138	Oxidation (M)
AIVDKVPSV	9	Q9Y678	
RVSEVKPKM(+15.99)	9	P33993	Oxidation (M)
KLLGLPEDYLY	11	P11388	
RVTQLTREEL	10	P12111	
RVSDFGGRSL	10	P27694	
NFYLVPLHY	9	P39210	
APRQPGLM(+15.99)A	9	Q9Y6H1:Q5T1J5	Oxidation (M)
EELNDLIQRF	10	Q15154	
EEVQVDPRKY	10	Q16659	
AVM(+15.99)NFVVRY	9	O60568	Oxidation (M)
EEVRQGLKAY	10	O76003	
M(+15.99)TDKAPPGV	9	Q7Z7J9	Oxidation (M)
RLLDLIPM(+15.99)SV	11	Q9NZ17	Oxidation (M)
RILDIDIPM(+15.99)SV	11	Q12800	Oxidation (M)
QVDPKKRISM(+15.99)	10	Q14680	Oxidation (M)

AFLNDSYLKY	10	Q8WVM7:Q8N3U4	
SM(+15.99)M(+15.99)DVDHQI	9	P48643	Oxidation (M)
ATLPRNFHY	9	O60716	
SFKSFEGLFY	10	Q00610	
YFGSNIAHM(+15.99)Y	10	Q9H3R0	Oxidation (M)
NPDIM(+15.99)RQTL	9	Q9UMX0:Q9UHD9	Oxidation (M)
NIANHFFTV	9	Q16222	
TLYEHNEL	9	Q9NRG9	
GEWASGGVRSF	11	P30044	
VLLDKIKNL	9	P12111	
GFSGHNLEY	9	Q9BUX1	
YLGRLAHEV	9	P40429	
SESLFVSNHAY	11	P04075	
SEQTGKELLY	10	P53396	
SLAETDKITL	10	P38606	
QIYDVGYQY	9	Q8IY17	
SEEELKKAY	9	Q5F1R6	
EELLDGHSY	10	P28070	
KLLSDPNYGV	10	Q9BTV4	
HM(+15.99)QNTPGFM(+15.99)Y	10	Q9NVP1	Oxidation (M)
EEYKQKIKEY	10	P62256	
SEVTGHSKGY	10	Q9HCJ3	
RENTQTTIKLF	11	P61978	
HFQHGGYQY	9	Q8NCA5	
EELQQKVSY	9	P40763	
SEIFHSQSF	9	Q9P0V9	
FYQRFPLSF	9	Q92611	
YLNEKAVSY	9	Q13492	
SEAPGPINF	9	P24844	
LEPEELYQTF	10	P13639	
SPHAVDGNTV	10	Q13151	
AIFSPM(+15.99)M(+15.99)Y	9	O43252	Oxidation (M)
VYTSVQAQY	9	Q6ZSR9	
TLIDLPGITKV	11	P50570:Q9UQ16	
TIFTARLYY	9	Q9NRW3	
AEINNIKI	9	P11388:Q02880	
QEIDEQRTKY	10	Q01850	
VRTPYTM(+15.99)SY	9	Q9BWF3	Oxidation (M)
TYAPVAFRY	9	A2A3N6:O60331	
GTLDSVLRV	9	O60287	
APQQTQHVSPM(+15.99)	10	P29692	Oxidation (M)

AEM(+15.99)LPGVLQQF	11	Q92597	Oxidation (M)
RPKHKEVHL	9	Q53EP0	
AIM(+15.99)DIVIKV	9	Q9NZJ5	Oxidation (M)
READQPLRTY	10	Q9Y4B6	
IPQGGRGAIQF	11	Q15437:Q15436	
RPTSKRDSL	9	P35251	
KYFGGVLEY	9	O94955	
APAPIHNQF	9	Q15011	
M(+15.99)LFGHPLL	9	P51784	Oxidation (M)
QLYEKGLYY	9	Q8NEZ3	
HTDITVKF	8	P11388:Q02880	
SVHKGFAFV	9	O60812:PODMR1:B7ZW38:B2RXH8:P07910	
EEIKKETGF	9	Q99653	
SPASIRTTDF	10	Q9H4H8	
KLEEIIHQI	9	Q15042	
AANPNRFITL	10	Q9Y4A5	
YQYPVIIHL	9	Q9BW60	
APRPDSRHSL	11	Q9BSV6	
M(+15.99)PYNHQHKY	9	O60427	Oxidation (M)
AESLIAKKI	9	P78371	
EENEKKQY	9	Q96NE9	
YFIDRDGEIFRY	12	Q96SI1	
NEIEDTFRQF	10	Q16864	
FYEFFNEQKY	10	Q14644	
SETGGSHDKRF	11	Q12906	
AFFEGPPFKF	10	O75083	
SPSIRNVQEL	10	P12109	
SPAKNPSSLF	10	Q9GZR7	
AESASEAAKKY	11	P51572	
LEEPTVIKKY	10	Q9Y5X3	
GENVIISHF	9	Q96AD5	
M(+15.99)VDGKPVNL	9	P63000:P60763	Oxidation (M)
NENSLFKSL	9	Q00610	
Y AidNPLHY	9	A5YKK6	
SPNKHNRLYM(+15.99)	10	P13639	Oxidation (M)
GPNPKSSRTTL	11	P17480	
KPAFFAEKL	9	P04083	
GLADASLLKKV	11	Q9UBB4	
ALLDRIVSV	9	Q92621	
HFEQFNINY	9	Q9HD67	
SLVPHNYGL	9	Q96TA1	

NELNIIHKF	9	Q8WWY3	
YITERIIAV	9	Q68CZ2	
SPIENIQRV	9	P35222	
AFDEIVALGY	10	P48444	
TLLEKQLEY	9	Q8IYX8	
STRYPPIAV	9	P12111	
APFLRNVEL	9	A5YKK6	
YSLEKVFGI	9	Q9UNF1	
QELQEINRVY	10	P07355	
QPHVGDYRQL	10	Q06323	
HNAEERRAL	9	Q07352	
AENISRVLY	9	P11216	
EEDSSGHTRAY	11	P52701	
RPKLPDNEV	9	Q7Z392	
SEVNRTGTQY	10	P50452	
HPKYPDGKTF	10	Q96SK2	
KAYN(+.98)QTHLY	9	Q14563	Deamidation (NQ)
VFVATFLRY	9	O00767	
FFSLPHVGY	9	Q969F9	
QPRKLGRQM(+15.99)	9	Q9P266	Oxidation (M)
FVDTRTLL	8	P08123	
SPQLREAVL	9	P12111	
YEM(+15.99)EVSQRL	9	Q86WA8	Oxidation (M)
VPDAKIRIF	9	P27635	
NFDKLSFLY	9	P53621	
GNYNDFGNY	9	P22626:P51991	
GILTKELLHSV	11	Q9Y217	
HFPATPLLDY	10	P53396	
ATSNVFAM(+15.99)F	9	P24844:P19105:O14950	Oxidation (M)
KEQNTQRSF	9	Q16665	
AETPDIKLF	9	P46782	
SEDHINRKY	9	Q9HAU0	
VAKAVTQAL	9	Q9Y490	
AVIENPEM(+15.99)LKY	11	P54198	Oxidation (M)
GETSYIRVY	9	Q13242	
KPQDYKKRL	9	P35222	
AEGTAQQRL	9	Q15149	
KIYEGQVEV	9	P46777	
TELERAFGY	10	P84103	
SYRDSYDSY	9	Q14011	
SEILDTTNSY	10	O75762	

EETVSGLKGY	10	Q10570	
SPNQKLLAV	9	Q9UNX4	
AEELFARKF	9	Q00610	
TPQSNRPVM(+15.99)	9	P24928	Oxidation (M)
SVFEKYFQF	9	O15144	
KVIEINPYL	9	P28074:P28062	
APVKPGPAL	9	Q9UBP4	
VPAEPKLAFL	9	P18124	
QM(+15.99)YDDYYYY	9	O60506	Oxidation (M)
IM(+15.99)KFINDQY	9	Q9UHD8	Oxidation (M)
APPRPGSSF	9	Q86XP3	
VLFEKEVNEV	10	Q9BZQ8	
SEHISSSGKKY	11	Q9BTA9	
GLIDRQVTV	9	Q9UPN3	
FQNALLVRY	9	P02768	
VAFKKTFYSY	9	Q99832	
ALFEEQGIGV	10	O95477	
HIKVTQEQY	9	P22681:Q13191	
LPKSPPYTAF	10	P23588	
RPFGSISRIYL	11	O75821	
QLAQFVHEV	9	A8MPP1:Q92771:Q96FC9	
RYFFTSVSRPGR	12	P04439	
YVYPKYLY	9	Q9BZ29	
PYNHQHEY	8	O95864	
SENPQGQTHF	9	Q9UKT7	
LM(+15.99)VDHVTEV	9	Q9HD15	Oxidation (M)
SPKQGSGEIL	10	P39059	
RPQPGRENF	9	P37802	
QM(+15.99)LEFAFRY	9	Q16594:Q9HBM6	Oxidation (M)
VPHKIITHL	9	Q13505	
RPSYHRQM(+15.99)	8	P41161	Oxidation (M)
SPVKSTTSI	9	Q9NQW6	
VAYGESHLF	9	P12109	
KVLEILHRV	9	Q5GLZ8	
QVLEPQLTY	9	P15559	
KLLSKFYEL	9	Q86UV5	
SPYPTKTQTM(+15.99)	10	P04843	Oxidation (M)
AATSRTYLY	9	Q8N0Z8	
NQFPGFKEV	9	P08579:P09012	
ATNASVFKY	9	P0DMP2:O75044:PODJJ0:Q7Z6B7	
NLLPKLHIV	9	O00299:Q9Y696	

EELEQKYTY	9	P42224	
NEIKDSVVAGF	11	P13639	
TTLDIITRY	9	Q86WA8	
RLDQRVPQI	9	P12111	
GM(+15.99)NNEVM(+15.99)KV	9	O75762	Oxidation (M)
TLSDLRVYL	9	Q9BYN0	
AVRALKLLTTL	11	Q32P28	
APSRSLNEL	9	Q8IY18	
RLQGHRQEV	9	Q9UM11	
AGELTEDEVER	11	P62269	
RPKGGRQINV	10	P12111	
NFGPKGFGY	9	Q16527	
AEDM(+15.99)ETKIKNY	11	P14854	Oxidation (M)
GLLTPQLHY	10	O95394	
TALHIYANY	9	Q96GQ5	
SFM(+15.99)GLLAM(+15.99)Y	9	O00391	Oxidation (M)
M(+15.99)VAKVGVSM(+15.99)	9	Q9P0S9	Oxidation (M)
ALQKRLDEV	9	Q9P2E9	
QIRPPSQDL	9	Q9UKM7	
HVSDHSGFHY	10	Q7Z6K3	
RPYTNKVITL	10	Q9NYV4:Q14004	
SLFEKGLKNV	10	Q9UKA1	
AEIEKLGQKY	10	Q9Y324	
KESERLVQY	9	Q8WUJ3	
M(+15.99)PSPVSPKL	9	Q9UHB7	Oxidation (M)
FGGPGTASRPS	11	P08670	
SLAKIYTEA	9	Q92522	
RLFDEPQLASL	11	Q9BX70:Q9H0C5	
GVILFHETLY	10	P04075	
SLSEKQYFL	9	Q8TCY9	
TPNKPSRTSM(+15.99)	10	P16615	Oxidation (M)
DPVKPPGSSL	10	Q13480	
ASKERSGVSL	10	P16402	
KIIGIM(+15.99)EEV	9	P52701	Oxidation (M)
ATVDSYHLY	9	Q6VMQ6	
VLYDRVLKY	9	Q9UHB9	
NQWDGTQHGV	10	P46939	
AVGEGFHNY	9	O00767	
EEAPVLM(+15.99)HY	9	P23219	Oxidation (M)
RPTDKPLRL	9	Q05639:P68104:Q5VTE0	
EENTNILKF	9	Q9NYL9	

QRQDIAFAY	9	P07355	
SEIEDHNRQL	10	Q9Y4B5	
NADAIIVKL	9	Q7KZF4	
TYTEIHQEY	9	Q96G28	
HLQAEVKQV	9	O75962	
LPYNQQHLY	9	Q9Y5Q0	
SYINFELRY	9	Q9BZJ0	
GLM(+15.99)KKAYEL	9	Q02078:Q06413	Oxidation (M)
GIM(+15.99)KKAYEL	9		Oxidation (M)
NPNLRKNVL	9	P23193	
QVDPLSALKY	10	Q9UL63	
SEEDLKVLF	9	P26599	
APAPAAHVA	9	P16989	
LPRQPPM(+15.99)SL	9	Q13045	Oxidation (M)
LLHDRSFSY	9	Q9BSJ2	
NEYGSINHTY	10	P11362:P21802	
GPESRPTL	9	Q9Y6J0	
NESEEVRFQ	9	P04632	
ILTDITKGV	9	P13639	
FIDASRLVY	9	P35221	
IIIDKEYYY	9	Q7L5N1	
AM(+15.99)PPPPQGV	10	O60216	Oxidation (M)
TGNKIRYNSK	10	P62891:Q59GN2	
REIESSPQY	9	Q8WWM7	
RPPVTKASSF	10	Q9Y2K5	
RLHGLPEQFLY	11	Q02880	
DESNALGKKF	10	Q6ZYL4	
KVADM(+15.99)ALHY	9	P50990	Oxidation (M)
YSQHM(+15.99)NHRY	9	O60315:P37275	Oxidation (M)
APDAKSFVL	9	P09382	
NVSRKAVRAL	11	E9PAV3:Q13765:O15069:Q9H009	
GTVEYDLRY	9	O00762	
EPSPPSRAL	9	Q13219	
KPRPIPM(+15.99)L	9	O94953	Oxidation (M)
SQWDSPM(+15.99)RV	9	Q7Z7G8	Oxidation (M)
REEHIDKVF	9	P40222	
GPGLSRTGV	9	P21333	
AIQDKLFQV	9	Q9BRQ6	
RQPDSGISSI	10	P04843	
LLDRFLATV	9	Q14094	
GPNSNRSQF	9	Q13356	

NADPQAVTM(+15.99)	9	Q9UNF1	Oxidation (M)
AM(+15.99)HTGSYGY	9	P09067	Oxidation (M)
SAALIIQKY	9	Q8IZT6	
SPAFSTRVL	9	Q7KZF4	
TLHKKTFYY	9	Q96D46	
SENAIVWKI	9	Q96CW1	
LPQGIVREL	9	P63272	
AEM(+15.99)YQVSRL	9	O94955	Oxidation (M)
IM(+15.99)YNLFQRY	9	Q9HD67	Oxidation (M)
YADVGGKQF	9	P51571	
QPNTTGKQL	9	P15311:P26038:P35241	
NEASVLHNL	9	P35579	
SPAKIHVF	8	Q7KZF4	
HPADSVVSVL	10	Q00325	
TENIDIKKY	9	O95749	
SPYQNIKIL	9	P52788	
SIDKTGENF	9	P62701	
GEVEQQRTY	9	Q15022	
EDKENYKKF	9	P08238	
AIDPHLLSV	10	Q29RF7	
EEAGRYLETY	10	P07992	
NPKESSSL	9	Q12789	
SGDGVTHTV	9	P60709:P63261	
APEEHPVL	8	P60709:P63261	
RPAEVGGM(+15.99)QL	10	P33316	Oxidation (M)
LPRPTQASF	10	P98082	
EELGHTETY	9	Q9Y2G9	
SEM(+15.99)DTSVKF	9	P06756	Oxidation (M)
YLHDFLKY	8	Q9UBU8:Q15014	
QQLDSKFLEQV	11	P42224	
GLFDQHFRL	9	Q9Y2I1	
M(+15.99)TFQM(+15.99)KFFY	9	Q15629:Q8N609	Oxidation (M)
AARPATSTL	9	Q04637	
NM(+15.99)SKIGYFY	9	Q969U7	Oxidation (M)
ADFGLARLY	9	Q9NYV4:Q14004	
ADFGLARIY	9	Q00534	
VPDGRILSV	9	Q8WUJ3	
KLFDKLELY	9	Q9BZZ5	
VIKETQERY	9	Q9H0E9	
TVADPSVISHLF	12	Q9NR19	
TPEEKLAAL	9	P40222	

RLM(+15.99)ELHFKY	9	Q8WYA6	Oxidation (M)
AEITGEIQQY	10	Q07889:Q07890	
KVLSKEFHL	9	Q01105	
GVLDELRY	9	P18583	
IVAIFPSPM(+15.99)LY	11	O95340	Oxidation (M)
TPASAGHVW	9	P48436	
AEPDNKNVY	9	P04844	
GLADKVYFL	9	P27708	
YIYPKHLKY	9	Q96BY6	
LM(+15.99)FTGHLLF	9	Q8IZU8	Oxidation (M)
SEAQVQKFL	9	Q9Y520	
YYPVVPVKY	9	Q14315	
APRQPPRHL	9	Q9P2P1	
EESQEIQKKL	10	O75116	
TPHQTFVRL	9	P04844	
AEHFLEQQM(+15.99)	9	P49368	Oxidation (M)
AVADKVHLM(+15.99)Y	10	Q14258	Oxidation (M)
HAISRYWYY	9	Q14669	
AEASM(+15.99)ISKL	9	Q13616	Oxidation (M)
RLAVYIDRV	9	P02545	
KLIEKLDIKL	10	P13639	
GPASRVSQL	9	Q9H7N4	
TRNDYVM(+15.99)M(+15.99)Y	9	Q96QU8	Oxidation (M)
NM(+15.99)NDLVSEY	9	P07437	Oxidation (M)
VVYYEPLVY	9	P48029	
QTLEHNFY	9	Q06730	
SFIDVDDERWHY	12	O14929	
ALKSSVEVY	9	Q8N4C8	
IARLPSSTL	9	Q86Y91	
KLIDDVHRL	9	P08240	
AVRKESGHSL	10	Q9Y512	
SIFHM(+15.99)SDSY	9	Q14690	Oxidation (M)
RPSSKKTYSL	10	Q96Q89	
NVTQVRAFY	9	P04040	
HPKETTRQL	9	Q15326	
SEEEIREAF	9	P0DP24:P0DP23:P0DP25	
SPALKTTRL	9	Q55W79	
GLDDIKDLKV	10	Q9HCE1	
APSQQRQQI	9	Q6P2Q9	
RLIPDTLYSV	10	Q99715	
TEFEDIKSGY	10	Q01105	

YRKPTYESY	9	Q04656	
YYTHPFQGY	9	P55265	
KPSSTRGGVLF	11	P39059	
FLLGPRLVLA	10	P49755	
RPLSKTVRF	9	P62280	
NFSDLVFTY	9	P35716	
ILENPPEEM(+15.99)KL	11	Q43432	Oxidation (M)
M(+15.99)PRQPSATRL	10	Q6P582:Q6NZ67	Oxidation (M)
KLGSVPVTV	9	Q9Y4C2	
VPRKPVKHTAM(+15.99)	11	P21333	Oxidation (M)
SPKATGVFTTL	11	P02751	
EELEKLQQY	9	Q9ULW0	
AEVFPLKVF	9	Q8NCS4	
AENPGKYNI	9	Q00839	
KLQELNYNL	9	P42224	
AEIDARNSDF	10	Q01082	
SYRDSYESY	9	Q96E39:P38159	
SELQARETF	9	Q15149	
SPHTPNEKF	9	Q32P28:Q8IVL5	
SSSSLEKSY	9	P60709:P63261:P68133:P68032:P63267:P62736	
IPRDPSQQEL	10	P49327	
M(+15.99)LLEIPYM(+15.99)AA	10	Q99613:B5ME19	Oxidation (M)
GPDKRLVTL	9	Q6EMK4	
RVDGPASNL	9	P12111	
SPTGSGKSL	9	Q9BX63	
YRGPSHTY	8	Q53EP0	
APWKPTWPA	9	O95084	
GPLFRNSQL	9	Q86YD1	
TPTRASPQL	9	P12111	
ATFSSSHRY	9	O94979	
APAYSRAL	8	P04792	
KM(+15.99)VGDVTGAQA	11	Q01628:P13164:Q01629	Oxidation (M)
NLAENISRV	9	P11216	
GVKKPTKAL	9	P54687	
SPKKPVIVF	9	O75643	
KPAAVRPPA	9	P12111	
SEIRHTADRW	10	P04792	
ALKDLINEA	9	P12004	
SEM(+15.99)PAAKRL	9	P35269	Oxidation (M)
KIYPLPHM(+15.99)Y	9	P21912	Oxidation (M)
M(+15.99)PAETIKEL	9	Q9H694	Oxidation (M)

NSSPLFLHY	9	Q63HR2	
QPAGEKRSSL	10	Q5TEA3	
RPRPSTDVRAL	11	Q96A09	
APYGNPKNM(+15.99)	9	Q92567	Oxidation (M)
LPHAPGVQM(+15.99)	9	Q13547:Q92769	Oxidation (M)
NM(+15.99)DHHILHV	9	P23219	Oxidation (M)
GAAGVHFIY	9	P52597	
EEIGSRNFF	9	P49023	
AEAEQTLRF	9	Q03001	
AEALPKQSV	9	P20290	
NGDFLPTL	9	A2A3N6	
SLDRPFTNL	9	P04844	
AEGIHTGQFVY	11	P62917	
KPKDPKVISEL	11	Q8IUR5	
AIFPSPM(+15.99)LY	9	O95340	Oxidation (M)
TAFGGFLKY	9	P25391	
KLDVGNAEV	9	P51572	
TLWGIQKEL	9	P00338	
KLGEFAKVLEL	11	P30153	
SPQGRVM(+15.99)TI	9	Q15365	Oxidation (M)
TLNEKLTAL	9	Q8WUW1	
NVGKTNFPY	9	P12111	
TLNEIQHQL	9	Q9Y6C2	
AIYHKYYHY	9	Q16394	
GEANGVWKY	9	Q6NUQ4	
NVDGHLIEL	9	P09936	
RPDSRHSL	9	Q9BSV6	
RPHHKILDM(+15.99)	9	Q08J23	Oxidation (M)
HIRDFVSRI	9	P12111	
TTFDKEFLL	9	O60306	
PDGHTKEM(+15.99)	8	P07814	Oxidation (M)
RLDDAQRLLY	11	Q6PK81:Q96HQ0	
SLTGRLDEV	9	P42226	
LLFEGEKITI	10	Q9H1A7:Q9GZM3	
AEFNRM(+15.99)SL	9	O43707	Oxidation (M)
YLRPPNTSL	9	O75494	
RPRPGTGLGR	10	Q9BVK6	
SILEHQIQV	9	P33991	
APDKTSTVL	9	Q9NR12	
GQVERFETV	9	P40189	
FLAEEGFYKF	10	P46977	

NPRKYLRV	9	P67809:P16989	
FPAGPPSHSL	10	Q96AY3	
SEAPSGESRKF	11	Q99613:B5ME19	
SPGPKRPGNTL	11	O75962	
NADTVSSKL	9	P63010	
SPHIPYKLL	9	Q9UPT9	
GTDGHVHLY	9	Q96EX3	
VLM(+15.99)TEDIKL	9	Q04637	Oxidation (M)
YYYDPTTGLY	10	P52756	
HVGM(+15.99)EPLQFDY	11	Q14671	Oxidation (M)
YLFHPLAY	8	Q9UKY4	
SPSSKYVKL	9	Q9H3F6	
LPAREATQF	9	Q14393	
APFHSHEL	8	Q68CZ2	
YYTKKEEQF	9	Q7L5N1	
VRPPVQVY	8	P48047	
KESEVFYEL	9	Q16665	
VLDDKLYVV	9	Q96PQ7	
EEVKLIKKM(+15.99)	9	P02792	Oxidation (M)
VSNPKSFEY	9	Q8IXI2	
RVSLVRKEL	9	Q96P16	
ALFNGAQKL	9	P61619:Q9H9S3	
HPDAPM(+15.99)SQV	9	Q9UBU8:Q15014	Oxidation (M)
TLAKYLM(+15.99)EL	9	P14635	Oxidation (M)
SPRYIFTM(+15.99)L	9	P11388:Q02880	Oxidation (M)
HYSQELSLLY	10	P78527	
SINAQRNAL	9	Q14669	
YYGHTNYIY	9	Q9Y263	
YM(+15.99)DYRGYSY	9	Q86WJ1	Oxidation (M)
APKRPPSAF	9	P26583:B2RPK0:P09429	
SEIEKNDVVF	10	Q99814	
AEGNSRFTY	9	P08123:P02452:P02458	
TLAEIAKVEL	10	Q15233	
KLKEFVDYY	9	O75151	
SPTKRLTL	8	Q12946	
NKADVILKY	9	O14602:P47813	
VPAKKITAA	9	P50914	
QLIDKVVQL	9	Q92503	
AVIEDTWHY	9	Q15800	
ALATLIHQV	9	Q9UBW8	
KIRPHIATL	9	Q14671	

ILDSQTEIYQY	11	Q96QB1	
FASHVSPEV	9	Q9NP61	
DSYVGDEAQSK	11	P60709:P63261:P68133:P68032:P63267:P62736	
RLHDVLM(+15.99)EL	9	P46939	Oxidation (M)
LPYKGGNTL	9	Q99715:Q05707	
KEIRDINRVY	10	P04083	
HFLDRHLVF	9	P60228	
SETDM(+15.99)VRQI	9	Q9UID3	Oxidation (M)
HYGNLGRLY	9	Q92624	
TPLASTSNSL	10	Q9NQW6	
LLRGPPVARA	10	P17936	
EEIKKNYRY	9	Q9BYV8	
NPSKKYNIL	9	Q9BUJ2	
APAQPRAAS	9	Q9UBX1:E7EW31	
QEDRALRSF	9	Q00341	
SLIGHLQTL	9	Q16690	
TPGGGSKTSL	10	P48681	
HEAEVLKQL	9	P16949	
APSQGHKSF	9	Q9Y4G2	
ILDDQTNKL	9	P98082	
GPRLPWTRL	9	Q6NUQ4	
TLFPGKVHSL	10	Q9NNW5	
EESLNIVKY	9	P86791:P86790	
YPQKIKAEI	9	Q8WYP3	
AEAPEHSF	9	Q15149	
KQADTTQYV	9	P16234	
GVSGIFM(+15.99)KY	9	Q96RQ1	Oxidation (M)
LLQATDFM(+15.99)SL	10	P04216	Oxidation (M)
AVKEKGSTL	9	O00767	
VLWDRTFSL	9	P42224	
KILDYEVTL	9	P40189	
ALLDKLYAL	9	Q9NV31	
EEAGGRPAM(+15.99)	9	Q00839	Oxidation (M)
GEM(+15.99)PPGLNF	9	P40855	Oxidation (M)
DEPLEGSEDR	10	P61978	
YVTTSTRTY	9	P08670	
FVDDQQTFF	9	Q16531	
SEKSEM(+15.99)QRHY	10	Q08117	Oxidation (M)
SPNTDREVL	9	Q14999	
SENNIQTIF	9	P05556	
GIADFLVKY	9	Q96EX2	

SLAQYLINV	9	Q15366	
TRDGQVINETSQ	12	P08670	
SADELVTRI	9	Q92575	
NPDAALRAL	9	O14579	
M(+15.99)THNLLLNY	9	Q13547:Q92769	Oxidation (M)
KPASTSGKSEL	11	O75592	
KM(+15.99)NEKLYTV	9	Q9NTI5	Oxidation (M)
KLDDTYIKA	9	Q99615	
KPKPVKPVTF	10	P51948	
AVM(+15.99)APRTLL	9	P04439	Oxidation (M)
QPTGGKARL	9	Q71UM5:P42677	
SVYGGGGIHEY	11	O75592	
NEVTTEIRF	9	P12111	
TLQKEIEKV	9	Q9BZF9	
AENISKDLY	9	Q9NQ31	
QPDQTRIVAL	10	P22087	
KQTSEYLRV	9	Q9H3M7	
SPM(+15.99)EVGKKL	9	Q9Y4G8	Oxidation (M)
AAAYPHTSL	9	Q9P2R6	
GTDELRLLY	9	Q9Y4W2	
SADNVRQQL	9	O43143	
SESSIDRLM(+15.99)	9	Q9Y678	Oxidation (M)
VLNDQHWY	9	P78357	
VEINIDHKF	9	Q00341	
RADDARQLF	9	P63096:P08754:P04899	
YM(+15.99)GHFDLLNY	10	Q9BWJ5	Oxidation (M)
ALFPGDVDRL	10	P11216	
GPAALRRAP	9	Q14393	
NVRPPNTAL	9	Q17RS7	
SPKQGSGEI	9	P39059	
SPQSTVTVF	9	Q16665	
RPSGKGIVEF	10	Q15233	
AEINENNVREF	11	Q9NS73	
SLTGHISTV	9	O43660	
VFLDSHFVY	9	Q99549	
FFGTHETAF	9	P51858	
SPHSGTISI	10	Q68CZ2	
ALNEKLVNL	9	O00303	
SADKPLSNM(+15.99)	9	P09874	Oxidation (M)
FTDEESRVF	9	Q9NYF8	
KLAPPSSTL	9	Q7Z3B3	

SLFPGKLEV	9	Q13045	
KLRDVTHSTM(+15.99)	10	Q99715	Oxidation (M)
M(+15.99)LYPLSHGF	9	O15294	Oxidation (M)
RLFPGSSFL	9	Q96BA8	
DENGKISRL	9	P62979	
NSDPNLHLL	9	Q96TA1	
SVM(+15.99)KTYHM(+15.99)Y	9	P0DMP2:O75044:PODJJO	Oxidation (M)
YYDPNSQYYY	10	P52756	
GFKSFEM(+15.99)FY	9	Q4LDE5	Oxidation (M)
ALAEKFEL	9	Q96EU6	
NPRGKIGLRL	10	P78357	
AVLGKGHQY	9	Q12769	
YVYEYPSRY	9	Q14511	
AM(+15.99)AASPHAV	9	Q13151	Oxidation (M)
RESEYKQVY	9	Q53EP0	
KVFQEPLFY	9	P07711	
M(+15.99)M(+15.99)N(+.98)GGHYTY	9	Q99497	Oxidation (M)
AEKLGGSAMI	10	P23528	
RVNNNRGNL	10	Q9Y5M8	
YYDGDVGNYY	11	Q13547	
YFVTGNHEY	9	Q6ZT21	
GLIDHQTYL	9	Q15149	
SEVDLTRSF	9	Q9P2T1	
KFSPLPLKY	9	Q9H2X6:Q86Z02	
ALNELLQHV	9	Q9Y490	
KEYDLAKKY	9	P49792	
TLTEEGVIKV	10	Q9BZE4	
KVQAGNSSL	9	Q9Y617	
YSDDIPHAL	9	P60228	
GPIYPGHGM(+15.99)	9	Q9UHA3	Oxidation (M)
KM(+15.99)KEIAEAY	9	P11142:PODMV8:PODMV9:P54652	Oxidation (M)
FVNDDRHVM(+15.99)	9	Q12906	Oxidation (M)
SEAAAYKEAF	9	Q04917	
RPSTSRSL	8	P08670	
SPASPKISL	9	Q8WWM7	
RPSGPSKAL	9	P48444	
GLELPEDEEEK	11	P08238	
YLLPAIVHI	9	P17844:Q92841	
KVLDIPSKY	9	P29375	
TENTEENRRFY	11	P04075	
NADQPLRV	8	Q99613:B5ME19	

FRTGNPTGTY	10	P17844	
SVITQVFHV	9	Q00341	
AESIVVHTY	9	P55060	
YLEEFITNI	9	Q9Y2D4	
YFIDSTNLKTHF	12	O00488	
KLYGKPIRV	9	Q15427	
RPAGEPYNRRTL	12	Q9ULL5	
QIIDAINKV	9	P12111	
EVYDDLHY	9	P53634	
KVFDGIPPPY	10	P40429	
KM(+15.99)DESKHEI	9	Q9Y490	Oxidation (M)
LVM(+15.99)APRTVL	9	P01889	Oxidation (M)
DFESAEDLEK	10	P19338	
LLFSHVDHIVA	11	P32418	
KPSSTTPRL	9	P27816	
M(+15.99)NYGTPISY	9	Q4LDE5	Oxidation (M)
ALM(+15.99)PVLNQV	9	Q5RKV6	Oxidation (M)
ILFDPKTSV	9	Q8N3C7	
SINDGFHVY	9	Q04771	
AEITDKLGLHSL	12	P61204:P84077	
TAADKRLTVL	10	Q9UHB7	
SLLPPDALVGL	11	Q15437	
ADGIVSKNF	9	P63220	
HLIHEVTKV	9	Q14697	
VHPPPTTTY	9	Q969N2	
YISDNVNKY	9	Q16665	
DEARDLIQRY	10	Q6P2Q9	
AEHLTSHKY	9	Q9NR56	
SPKGKFSLF	9	Q09666	
KPGDFIHTL	9	P04818	
KLHYVVTEV	9	Q9Y4A5	
SQADIKKAY	9	Q9Y2G8	
SPTKKVAV	8	P19338	
HAANQSYQY	9	O43707	
TIDEIQKL	8	Q16531	
ADYQQTIKKY	10	Q14203	
YFFTSVSRPGR	11	P04439	
AESLQPKGY	9	Q6ZRS2	
KEAETRAEF	9	P07951:P67936	
KQM(+15.99)EQVAQFL	10	Q01995	Oxidation (M)
KPHPSPTSL	9	Q7L9B9	

APRTIFQRV	9	Q9UBB4	
HPRPNFRQM(+15.99)	9	Q14315	Oxidation (M)
TPSSKSKEL	9	Q04727	
SLFRVITEV	9	Q93063	
SWNNHSYLY	9	P38435	
GPHPPAAHL	9	Q13207	
RQFDHPHIV	9	Q05397	
HPHSGSGRSVL	11	Q12772	
SEIEAKVRY	9	Q9Y490	
ALASHLIEA	9	Q9NZN4	
NVM(+15.99)GEQFSY	9	Q15904	Oxidation (M)
EELGFRPEY	9	P09874	
SPRVPNSSV	9	Q9NR09	
AESPVFKEL	9	Q08357	
APGIRDHESL	10	P51858	
APRPPKPM(+15.99)	9	Q5JNZ5:P62854	Oxidation (M)
RLDELGGVYL	10	P04844	
YLANGQTKV	9	Q9H3M7	
RSPSFTHF	9	P23219	
YLDQISRY	9	P61289	
QLYEALTHV	9	O76074	
KRNDYVHAL	9	Q99873	
QLVDIIEKV	9	P61289	
HEAQDAGYRM(+15.99)Y	11	Q08209:P48454:P16298	Oxidation (M)
EVVEEAENGR	10	P06454	
RLYQGINQL	9	P43246	
AEAECTFSY	9	Q9Y296	
SPSKPAKSL	9	P86791:P86790	
QVYGNPQQY	9	Q01085	
EVM(+15.99)DVFLRF	9	Q9P2E5	Oxidation (M)
YVPDVTGRY	9	P21333	
VPQKTIQEM(+15.99)	9	O43707	Oxidation (M)
YLDPRITVA	9	P11387	
AVRELDGRTL	10	P84103	
SLSTFQQM(+15.99)W	9	P60709:P63261:P68133:P68032:P63267:P62736	Oxidation (M)
KLVDKSTEL	9	P12111	
APQRSQSPL	9	O95817	
YLGPHIASV	9	Q96GC9	
HTDLPVSFV	9	Q7L5N1	
QLRLPGQQL	9	Q15036	
VLFENTDSVHL	11	P42696	

AEQESARQL	9	Q15149	
M(+15.99)PKKDARTL	9	P29083	Oxidation (M)
EEVYAGSHQY	10	Q9H3P7	
YLPDM(+15.99)SGRY	9	Q14315	Oxidation (M)
RPQRATSNVF	10	P24844:P19105:O14950	
SPM(+15.99)ADRSM(+15.99)M(+15.99)	9	P18583	Oxidation (M)
AELERQQIL	9	Q8WWI1	
SPGARLITI	9	P21964	
APQPGKYVI	9	P21333	
SPSTNRGPIQM(+15.99)	11	Q6VMQ6	Oxidation (M)
TENPVIM(+15.99)GL	9	P26196	Oxidation (M)
GLFPYPYTY	9	Q13207	
EENKLVKKI	9	Q16181	
NELVSTHGY	9	Q9UM11	
SADPQHVVV	9	Q63HR2	
RTLNHQYDY	9	P48729:Q8N752	
AEM(+15.99)AQNAVRL	10	P35222	Oxidation (M)
SEQPAKATF	9	P49790	
SNLGSAYHY	9	Q96AY4	
SLLHLGALY	9	O00767	
AM(+15.99)LTVLHEI	9	Q8N3C0	Oxidation (M)
AELESSKEQF	10	P52306	
SVDVTNTTF	9	O14579	
HPEKPKKEL	9	P16234	
VQTEVNREL	9	Q9H2G9	
SPSQEIHARF	10	P00750	
LPDENREVL	9	Q96QB1	
NPM(+15.99)VTGTSTV	9	P28070	Oxidation (M)
ASDPDEAKL	9	Q13045	
ALSKEGIVAL	10	P40227	
RTLPLVLLLY	9	Q9NR09	
NPDM(+15.99)RVQIL	9	P53396	Oxidation (M)
IPDKPRHIQM(+15.99)	10	Q9NW15	Oxidation (M)
KLLDTM(+15.99)VDTFL	11	Q6P1K2	Oxidation (M)
EEM(+15.99)RKLFEKY	10	Q15233	Oxidation (M)
GSFHGDFAY	9	O15027	
QVM(+15.99)PQLLNY	9	Q16342	Oxidation (M)
SPDKRLTL	8	O43524	
APYPLYPAEL	10	Q93062	
KILEELQKV	9	O75179	
ASDKETYEL	9	Q6P5R6	

RAAPVTGGM(+15.99)	9	P18583	Oxidation (M)
HPIYNIKSL	9	Q13601	
M(+15.99)ENIHSEM(+15.99)Y	9	P31350	Oxidation (M)
GLSTEGIYRV	10	Q9NRY4	
QQIGKVVQV	9	Q9UNX3:P61254	
QNSPSTFHY	9	P21359	
RNFDKAM(+15.99)SY	9	Q96AY4	Oxidation (M)
APAPSTVHF	9	Q13118	
TLIEQQFTY	9	Q8TCG1	
AVRPPAAAA	9	P12111	
WIGDKSFEY	9	P58317	
AVAEKM(+15.99)LQY	9	Q9NSY2	Oxidation (M)
RESKKIDRHM(+15.99)Y	11	P84098	Oxidation (M)
ALAEIAKAEI	10	P23246	
SESGAGLTRF	10	Q9UNF1	
EESGARINI	9	Q15365:Q15366	
SLAEGRLYL	9	Q7Z3E5	
FYSEVHPHY	9	Q9BRR6	
NRLEATSY	9	O60341	
KIADRFLY	9	P61968	
AEEKLTHKM(+15.99)	9	P16949	Oxidation (M)
YLTNEGIAHL	10	Q15149	
SPGQRSISL	9	P00519	
ILDKKVEKV	9	P08238	
KPTGTPKAL	9	Q99685	
AIDFPESKL	9	Q07954	
IVDKPAEF	8	Q14315	
KLNDLIQRL	9	P53621	
TPFKGGTL	8	P42166	
GVM(+15.99)PPTQSF	9	Q9UJM3	Oxidation (M)
AETVQTVRY	9	Q8N0X7	
SFARLSLTY	9	Q15904	
AVYTIALRY	9	P78382	
VTFPEFLRY	9	Q8NCH0	
HPRVPSLTTV	10	Q9BW19	
QVLDRYHRF	9	Q13813	
TEM(+15.99)EGQAKEL	10	Q9BZF9	Oxidation (M)
SLKDM(+15.99)VFLDY	10	Q9BYM8	Oxidation (M)
TENQRPKTGF	10	O75717	
SYPPRGFSL	9	Q9Y6C2	
GIHTGQFVY	9	P62917	

LYYDPNSQYYY	11	P52756	
VLIDYQRNV	9	O14980	
RPPTAGSQF	9	O43795	
KKSDPVVSY	9	P13639	
TPAPPTHAL	9	Q5SXM2	
TLTSKLYSL	9	Q9UDW1	
NRGPAYGL	8	P37802	
SLFAGGM(+15.99)LRV	10	Q9BVK6	Oxidation (M)
KPKAPPSL	9	Q9NZZ3	
AEQM(+15.99)PQHTL	9	Q3L8U1	Oxidation (M)
AEHDRIARQW	10	Q969T4	
GADNSVVRF	9	P21333	
NPQERTLTL	9	P08238	
NVADYYPEY	9	P61421	
AAAAGRIAI	9	P26599	
SPTGTGKSL	9	A8MPP1:Q92771:Q96FC9	
SLYDRPASY	9	P16234	
QQQQPQHAF	9	Q99967	
KPSEERKTI	9	Q14865	
VDFESAEDLEK	11	P19338	
SVLDSFLKY	9	Q96JI7	
KPRSPDHVL	9	Q9UPN3	
YEKNPDSHGY	10	Q9NX14	
RLFDEPQLA	9	Q9BX70:Q9H0C5	
SVM(+15.99)RLTISY	9	Q16665	Oxidation (M)
EDNGIIKAF	9	P36578	
APIQGSRL	9	Q16665	
TEVDSRWQEY	10	Q9UPN3	
RPHKVDGRV	10	P09651:P51991	
RNLDKAFSY	9	Q9Y6R0	
SLYSQVHQI	9	Q93074	
GGPGTASRPS	10	P08670	
KLWEAESKL	9	Q9NYP9	
SPM(+15.99)AERSM(+15.99)M(+15.99)	9	P18583	Oxidation (M)
APYSRPKQL	9	Q86V81	
KTASINQNV	9	Q8NI77	
FLM(+15.99)DFIHQV	9	Q8WUM0	Oxidation (M)
SEEEFLRTY	9	Q9NSA3	
QEGDTRKY	9	Q15596	
RASPIKM(+15.99)DL	9	P27816	Oxidation (M)
IVDDITYNV	9	P02751	

KESDGFHRF	9	O00469	
SLDQPTQTV	9	Q99613:B5ME19	
SVLDQKILL	9	O60287	
YYDGDIGNYYY	11	Q92769	
EETYSPIKIF	9	P24593	
SPSRQLVGRTF	11	Q86UU1	
KLIGQVHEV	9	Q5TBA9	
KLLPGDLTQV	10	Q9H7D0	
HENEDIYKL	9	O00505	
APSPRPLSL	9	Q6NUT3	
SETPEIRKY	9	Q8N4C8	
STYEKALGY	9	Q96DB5	
KYPPVNRKY	9	Q8IWT0	
SLAQKKDRVAQ	11	P46777	
SLDVGQDRV	9	P12111	
HVNDLFLQY	9	Q68CZ1	
AM(+15.99)AARPHSI	9	P22626	Oxidation (M)
GLNEEIARV	9	O14777	
TEDGKVEKHY	10	O75436	
YILDQTNVY	9	Q9BUJ2	
FRVGNVQEL	9	P12111	
RPEEGRPVV	9	P27816	
RLPDDPTAV	10	Q9UBF2	
AEGPPRLAI	9	Q15773	
ALDESHNQNL	10	O94804	
EEYPDRIM(+15.99)NTF	11	P07437	Oxidation (M)
M(+15.99)EFERQAEI	9	O43633	Oxidation (M)
AM(+15.99)ASHLTST	9	Q9BWF3	Oxidation (M)
AEIDDRKQL	9	Q96P16	
SVKPASSSF	9	Q9Y4W2	
KVFDPPVPGV	10	Q08211	
RM(+15.99)LPHAPGV	9	Q13547:Q92769	Oxidation (M)
ALIEDEHNL	9	Q49A88	
APRLPITGL	9	Q8TF50	
APRPASGP	8	P26373	
NLLTHIENV	9	A8MW92	
PRPPPKPM(+15.99)	8	Q5JNZ5:P62854	Oxidation (M)
GM(+15.99)NQQAYAY	9	P17844	Oxidation (M)
RPDSRLGKTEL	11	Q9BYW2	
FRQDHPSSM(+15.99)	9	Q01844	Oxidation (M)
AGDDAPRAV	9	P60709:P63261:P68133:P68032:P63267:P62736	

APASFGKSF	9	Q96EY5	
RNFHVFLEY	9	Q9Y5X3	
QELLM(+15.99)AHDF	9	Q9NZN4	Oxidation (M)
YLFERIKEL	9	Q96SK2	
KILPTLEAV	9	Q12905	
AADDHQALI	9	P61962	
SEQSLGQKL	9	P33991	
SM(+15.99)LEDVQRA	9	Q9NW13	Oxidation (M)
M(+15.99)PTGGGKSL	9	P46063:P54132	Oxidation (M)
KLLEKLYKV	9	Q15034	
KLFGM(+15.99)IITI	9	P61619:Q9H9S3	Oxidation (M)
NLYPFVKTV	9	P31939	
LPDQDQRSQL	10	Q13889	
RPERVKPFM(+15.99)	9	P13693	Oxidation (M)
ALDKATVLL	9	Q53EL6	
EEASLLHQF	9	Q01082	
NPNLRIISL	9	Q92484	
LDDGIDDER	9	P11940:Q9H361	
SPQKGRGGLGK	11	O43680	
IPRITIEM(+15.99)	8	P11388	Oxidation (M)
GENDRGVSF	9	P62136:P62140	
GLWGPVHEL	9	Q9Y4C2	
TYEAVREV	9	P62906	
SPGPPTRKL	9	Q9P2D3:Q86XA9	
KENSITKY	9	O95391	
SPRGGTPGSGAA	12	O14497	
GLIDYDTGIRL	11	Q03001	
RELLPVKKF	9	Q08211	
IPLSKIKTL	9	P13010	
TVDGPRNEL	9	P55265	
TM(+15.99)M(+15.99)SRPPVL	9	Q5T9L3	Oxidation (M)
AEIEIM(+15.99)KKL	9	Q9Y6M1	Oxidation (M)
TENKERKSF	9	Q9P0V3	
ATLKDQLIY	9	P00338	
SLKTPGTSL	9	Q76L83	
EEQRAIAKI	9	P78344	
AESYRIQTY	9	O95453	
TPNKAGPKL	9	Q9UKJ3	
KLYDGFQYL	9	Q9UBF8	
VYPQYVIEY	9	Q7Z2W4	
SPVVRAENL	9	P41252	

AADKRLTVL	9	Q9UHB7	
APARLFAL	8	P31431	
SEGPPRLW	9	Q96QB1	
EEGDRVITY	9	Q13619	
KPTDEKLREL	10	Q14683	
APSKTSLIM(+15.99)	9	Q9H334	Oxidation (M)
RESPSVKRM(+15.99)	9	Q02078:Q06413	Oxidation (M)
QLFSQQHFL	9	O15417	
RPHIPISKL	9	Q8TF42	
RPRHQGVM(+15.99)V	9	P60709:P63261:P68133:P68032:P63267:P62736	Oxidation (M)
KLFEEIREI	9	Q8N4U5	
SM(+15.99)PDFDLHL	9	Q09666	Oxidation (M)
APEEHPTL	8	P68133:P68032:P63267:P62736	
DEALIGKKF	9	O14647	
AVRIPGGPA	9	P09958	
RPTAAPRSM(+15.99)	9	Q9NQW6	Oxidation (M)
RLDEDGNTAL	10	O75762	
KIADFGWSV	9	O14965	
GPM(+15.99)PKHIAF	9	Q86SQ9	Oxidation (M)
NELQSIQRL	9	Q13207	
TLADVLYHV	9	Q9UBL3	
TVTEVLLKY	9	A6NIE6:Q9NYV6	
RPSGPGPEL	9	P51531	
RETGSSKGF	9	P19338	
SESTNQRVL	9	Q9BVK6	
ILDQKINEV	9	P11926	
QVFPGLLERV	10	Q9UBB4	
ALSPNNHEV	9	Q92747	
SFSQEVSHM(+15.99)Y	10	O15403	Oxidation (M)
NPRLNIDTL	10	Q13200	
SEHGAQHHVY	10	Q99594	
SRKESYSVY	9	Q99879:Q93079:Q99877:O60814:P62807:P58876:Q5QNW6	
AEQDFISKF	9	Q9NV70	
TELPKAEKW	9	Q9ULR0	
QVHNLLTY	9	Q15041	
KPYTSFQEL	9	P23219	
FYLGGSLSY	10	P55083	
GLLERVKEL	9	Q00341	
EEQRGLRKY	9	Q9BZV1	
TEM(+15.99)ESQKQL	9	P78316	Oxidation (M)

RLYPEGLAQL	10	P61421	
HPNGNITEL	9	Q7KZF4	
VVFEIQGTTV	10	O94955	
VLIEKSTEM(+15.99)	9	Q9BZF9	Oxidation (M)
SEM(+15.99)AGPPQM(+15.99)	9	Q6P2Q9	Oxidation (M)
KIFDEILVNA	10	P11388:Q02880	
KLQEFLQTL	9	Q9NVI1	
QLFSEIHNL	9	Q96SN8	
EEEHGKGREY	10	O14646	
NELDAVQREY	10	O14777	
NPVHNGHAL	9	O43252:O95340	
SPPLRPRL	9	O75882	
YFVKFLDAY	9	P04062	
KEIGDVENW	9	P78537	
AETPQKPPY	9	Q12952	
STNAYM(+15.99)LIY	9	Q96K76	Oxidation (M)
SENELKKAY	9	O60884	
GEDPAQPRKY	10	Q12873	
SEPDFVAKF	9	Q9NWX9	
APGSARRPL	9	P11279	
GPAHPTNNF	9	Q8IY63	
SPRVPASET	9	Q9HD15	
SFLPSESGPYGY	12	P05412	
QALDIRLSY	9	Q5THJ4	
KIIDFIYTA	9	Q9P2J3:Q9P2N7	
KEYERAIF	9	P27105	
TPRHIVRF	9	Q9UN81	
KM(+15.99)PEINAKV	9	Q15011	Oxidation (M)
RLDDNIRTV	9	Q5VIR6	
AILETAPKEV	10	Q9P2E9	
TEQDLKEYF	9	Q13148	
APRKPDWDL	9	Q8WUD4	
EEVDLSKDIQHW	12	P31350	
IEDGVPQHL	9	P12111	
REAAATKRL	9	P04183	
AELKQVQTL	9	Q9UPN3	
APARLFALL	9	P31431	
RPRSALPRL	10	Q14332	
APAPSSKDL	9	Q14774	
ILLDHEKEWKL	11	P54136	
LLDPNVKSIFV	11	O43414	

FLDGNEM(+15.99)TL	9	Q9Y696	Oxidation (M)
SQDFPGSNL	9	P15036	
SIIGRLLEV	9	P62136	
QLYEFDIKV	9	P15586	
AEHPELSY	9	Q12765	
GARFPSGTL	9	Q13443	
IIGITVLVVAVP	12	P23634	
SPNRSFTEM(+15.99)	9	Q9Y263	Oxidation (M)
RLQEENQQL	9	Q8N283	
VM(+15.99)DPKQM(+15.99)NV	9	Q9Y646	Oxidation (M)
EELKELKEY	9	Q9GZU8	
KPSEERVREI	10	P31350	
KLDPTKTTL	9	Q92597	
KQVEEIFNL	9	P78347	
AELTGDHQL	9	O75643	
ALNEQIARL	9	O14777	
SEKQFSKRY	9	Q6P5R6	
TPRVPDNL	9	Q13829	
SLLSHVEQL	9	Q9UI95	
LVRQQLSEM(+15.99)	9	Q9P2E9	Oxidation (M)
YYYGQIHM(+15.99)Y	9	P10070	Oxidation (M)
PGRPGLTL	8	B3EWF7	
RPSQPTQET	9	Q9H3M7	
NADSATRL	9	P27105	
SVAGKIHTV	9	Q9Y6Q2	
SSDGHEFIV	9	Q15369	
YLHDQFWSY	9	Q9Y2V0	
ARPPGYEF	8	Q8IU81	
NM(+15.99)YEGVGRM(+15.99)FI	11	O60925	Oxidation (M)
NTVTRQWLY	9	Q9P2K6	
RIM(+15.99)EKIRNV	9	Q08AF3	Oxidation (M)
GPAKWQNVSL	10	Q9P266	
AEKAPTIVY	10	P15927	
LPKQPPLAL	9	Q96S59	
VPRGQGSQF	9	O75807	
HARPATSTV	9	Q5T0B9	
ALRDVSEEL	9	P40222	
EEFDAITKAY	10	Q9ULC3	
AEM(+15.99)IRQIEF	9	Q7Z3E5	Oxidation (M)
NPRPTAQDL	9	Q9Y2H0	
AVIDVGINRV	10	P13995	

GPGLNRTGV	9	Q14315	
KLLPQLTYL	9	Q92688	
KEAEFLQKL	9	A2A3N6:O60331	
RPKQPGPVKL	10	Q9Y618	
ITAPPSRVL	9	O00767	
KTFPYQHRY	9	Q8NI27	
VADKFTEL	8	P30153	
SPTKLDVTL	9	P27816	
SEIAQKQKL	9	Q9NVD7	
AETEGYLR	9	O60568	
KPNNPKHTL	9	Q9UBT6	
AFNSKQLLY	9	Q12974:Q93096	
SEAEQTLRF	9	Q9UPN3	
KEIKDVKEM(+15.99)	9	P98082	Oxidation (M)
VGDIPGVRF	9	P62266	
KVDQEVKL	8	P61289	
GLDPNKPEL	10	Q9ULL5	
NVFIQGFY	9	Q2M389	
SYLPRSFY	9	Q08AG5	
RLNEKNYEL	9	Q9UI12	
LPSSGRSSL	9	O95817	
YYKPDSPEY	9	Q86U86	
NEGTGKSSW	9	O43524	
DEITGKKGY	9	Q14149	
RPQGGQDIL	9	Q9Y265	
VPLGPFLGK	9	P16930	
KEYERFSKY	9	Q13838:O00148	
AAVRIGSVL	9	Q9Y5B0	
KPHSGFHVAF	10	Q8IYU8	
TEAETIQKL	9	O60763	
SEGLLRLY	9	Q12789	
NLDNPIQTV	9	P69849:Q15155:Q5JPE7	
KISTITPQI	9	P30260	
SEHQPTTW	8	P46977	
AM(+15.99)FENFVSV	9	Q09161	Oxidation (M)
HNSPAYLAY	9	Q969G3	
VM(+15.99)TSLQQEY	9	Q05397	Oxidation (M)
LM(+15.99)EHIHKL	8	P84098	Oxidation (M)
APAPRPSLL	9	Q92620	
SPSKRLTL	8	Q12947	
YVIQKFFEF	9	Q14671:Q8TB72	

QLLGAHEV	9	Q13813	
APAERQIDL	9	Q29RF7	
KVRYPDRTL	10	P60510	
TARPPGYEF	9	Q8IU81	
RLQAQRNEL	9	P62195	
REAKPHM(+15.99)ESY	10	Q8IVL5	Oxidation (M)
RPSSRSYVTT	11	P08670	
FVHDLVLYL	9	Q00610	
SPSTSRTPLL	10	P00533	
TELDLHKRL	9	Q9P2N5	
APAPPKAEA	9	P62750	
LLDEKVQSV	9	Q63HK5	
SELPRFQEL	9	Q96TA1	
SEILQKEVY	9	Q9NRW7	
LPKATILDL	9	Q96AG4	
QEIYGGSDSRF	11	O75533	
VTDRVIGLL	10	P12111	
GLWKSETTV	9	Q96PE1	
EPRPPHGEL	9	P04818	
SLAEGSVTSV	10	P13010	
EESEAYKSL	9	Q9BY42	
RPPPIGAEV	9	Q07820	
KIAPTVWKL	9	O00391	
GA AVRIGSVL	10	Q9Y5B0	
NQDRPSLL	8	Q96G23	
NRPPFGQGY	9	Q92734	
SLFHAQLAY	9	Q16537	
VILEGELERA	10	P67936	
KEEQFKQVF	9	Q7L5N1	
RLFKDQLVY	9	Q9Y3I1	
PRPSTGPHKL	10	P62701	
FLLDPVKGERL	11	Q15149	
RAREEGIRL	9	P12110	
IQVTKVTQV	9	P48444	
RPAANPIQF	9	P23921	
FLKDLVASV	9	Q14919	
APGVRABEL	9	Q9Y508	
GVRPVSGDL	9	Q9Y2U8	
DTLRDPGSQL	10	P11717	
HPRSPNVL	8	Q16665	
LPSLRILYM(+15.99)	9	P00403	Oxidation (M)

GLIEKNIEL	9	P26358	
DEVERVITI	9	P62269	
NPGAKRILEL	10	P12956	
ILM(+15.99)GVLKEV	9	O43143	Oxidation (M)
GTIGPGKYGY	10	Q6ZMP0	
VPNQKRLTLL	10	O60488	
RPDRGRLGFQV	11	Q01995	
LLYQEGAKM(+15.99)AV	11	P14923	Oxidation (M)
SLHDIVGNLV	10	P12111	
RVAEHHVATL	10	Q12834	
APLSGSKEM(+15.99)	9	O95391	Oxidation (M)
NPM(+15.99)VTGTSVL	10	P28070	Oxidation (M)
M(+15.99)PPSGPRGTL	10	Q02818	Oxidation (M)
GLQEGTHEL	9	O76094	
AELEDRRQL	9	Q9NQG5	
VLAKPGVISV	10	P12111	
SPNSKVNTL	9	P40939	
SPSHPANSF	9	Q9BUD6	
RIREHVPQL	9	P12111	
AESPAGHSF	9	Q9NXE4	
M(+15.99)PRTGNGPM(+15.99)SV	11	O95793	Oxidation (M)
HVLEEVQQV	9	Q07065	
SPAHRPALLL	10	P78539	
SYLPDM(+15.99)SGRY	10	Q14315	Oxidation (M)
SPGGPRAAV	9	P53814	
KPM(+15.99)SYEEKRQL	11	O60885	Oxidation (M)
YRNDYATM(+15.99)L	9	P12110	Oxidation (M)
GEFKALQEW	9	P24928	
KVLGIVVGV	9	A5YKK6	
KPRPPPPPP	9	Q8N9Z2	
RLIESLFTI	9	Q12789	
GVAERSWLY	9	O00255	
AEKNVEKKI	9	P40429	
AVHDITVAY	9	Q6UWP7	
SEHPLAQLY	9	O76074	
ATDPTIHTV	9	Q9H6U8	
KPRDVSSV	8	Q01082	
TTSHLM(+15.99)GM(+15.99)FY	10	P39748	Oxidation (M)
SEM(+15.99)RFVQDL	9	O43293	Oxidation (M)
GLDIDGIYRV	10	Q8IWW6	
SEESFKQYF	9	Q6DKJ4	

RPRGPPPPP	9	Q13435	
QPHQVEQTEL	10	Q14444	
KLDQDLNEV	9	Q9UQE7	
QLLEKVIEL	9	Q9NTX5	
AADGKGVVV	9	P46779	
RPQEGGRITL	10	Q03001	
NPAKVWTDL	9	Q5SRE5	
VPNNKITVV	9	P07195	
SM(+15.99)FNGQLLF	9	P52272	Oxidation (M)
SPNLRLLDL	9	Q8N531	
KLQDGVHII	9	Q9Y3A2	
RPKTTKSGKAL	11	Q96JG8	
RLWGEPVNL	9	Q93008:O00507	
AGDGGRVAL	9	P30559	
SSHSNAYFY	9	O75844	
SPQKGRGGL	9	O43680	
RELQHEHSL	9	Q14517	
SEALPGDNVGF	11	Q05639:P68104:Q5VTE0	
KTTTIAVEV	9	Q00341	
LM(+15.99)TKEISSV	9	P78527	Oxidation (M)
FLDEKTHEL	9	Q9H0K6	
QVRLPLGTL	9	O43148	
ALREPGKAL	9	Q8IU81	
FLYKIIDEL	9	P12111	
KLHDVEHVL	9	Q99471	
RLAEDVQRA	9	Q9Y6C2	
APAPVGPLVGL	11	P98082	
KQVSDLISV	9	P17844	
FLDNERHEV	9	Q16512	
TLPRNFHY	8	O60716	
GPAPAPHSL	9	P41161	
EEESQPAKKTY	11	O75400	
DEKDEHGRAY	10	Q1KMD3	
RAAPLQGM(+15.99)L	9	P12111	Oxidation (M)
ALGGHPLLGV	10	O94907	
KAYNQTHLY	9	Q14563	
VPATDRNAL	9	P60228	
SADFETVRM(+15.99)	9	P61160	Oxidation (M)
REHDIAIKF	9	P30260	
KEALPKAAF	9	Q13123	
VM(+15.99)DRGYSY	8	P28074	Oxidation (M)

GNLDRAYEF	9	Q00610	
AEVAAEKSF	9	Q9NW64	
RPQKSSTRL	10	Q14517	
VM(+15.99)FKEPVEV	9	Q9BX70	Oxidation (M)
ALLKVNQEL	9	Q9Y512	
YAFGKYLGY	9	P21589	
IM(+15.99)FGVPVYV	9	P18074	Oxidation (M)
AYPEYLITY	9	Q9H2K2:O95271	
ALLDSQFSY	9	O15372	
IARNLTQQL	9	Q99541	
SPQEFIHEL	9	Q9NSV4	
SEIDKRVQY	9	Q9HCG8	
RPKKPGQSF	9	O15042	
SPQEARNM(+15.99)L	9	P33240	Oxidation (M)
TSVSRPGR	8	P04439:P01889	
SPM(+15.99)GVDHM(+15.99)SGL	11	Q15717	Oxidation (M)
AEAERLLLI	9	Q9NQW6	
FSDTPVTEF	9	P12111	
PRPDSRHHSL	10	Q9BSV6	
VSDFGGRSL	9	P27694	
SPQPRLIYL	9	Q96A54	
YYIEYNFKY	9	O75762	
KILPDLNTV	9	O60287	
ATADLHAQY	9	Q8NOW4:Q8NFZ3	
KPANVFITA	9	Q9HC98	
ALAEQQEKM(+15.99)LI	11	O76074	Oxidation (M)
APASPRQL	9	Q07021	
KPSSWLRTL	9	P12755	
AERDVYTGDAL	11	P20618	
AELELQKQL	9	O43290	
RPLEM(+15.99)IEPRTL	11	P62249	Oxidation (M)
LPQKPHLKL	9	Q6VMQ6	
QADFEAHNI	9	P49915	
ALANHLIKV	9	Q9H4M9	
NPRDPPSWSV	10	Q9Y490	
SEPSEVRSY	9	Q15149	
SPREALIRL	9	O95125	
TPM(+15.99)GPGRTV	9	O00214	Oxidation (M)
ALFDGDPHL	9	Q8WUY3	
RIRM(+15.99)TDGRTL	10	Q9BRA0	Oxidation (M)
TM(+15.99)DPKPLAY	9	Q7Z3F1	Oxidation (M)

EVVEEAEN(+.98)GR	10	P06454	Deamidation (NQ)
EQLEEEEEAK	10	P35579	
FRQPVVDTF	9	Q86X55	
KLFNEFIQL	9	Q9UNX4	
TTFEHAHNM(+15.99)RV	11	P13639	Oxidation (M)
AETDFEKRDF	10	Q99615	
ISDLGQHTL	9	P39656	
RIAKEKKAL	9	Q96L91	
LPSDLRTISL	10	Q9C0A6	
GQLVKM(+15.99)LLY	9	P50395:P31150	Oxidation (M)
QLLDFERTL	9	Q16829	
ILLDNDHYAM(+15.99)	10	Q86WA8	Oxidation (M)
QPSAPGHSV	9	Q9ULH7	
GPNTNGSQF	9	P62937	
RVRNKLIVM(+15.99)	9	Q9UHQ7	Oxidation (M)
APYRDLPASL	10	P54259	
KVIDTQQKV	9	O60925	
AVQNIRNVL	9	Q9BVR0:O95714	
SLQEEIAFL	9	P08670	
SPRPAAM(+15.99)VL	9	Q9HB63	Oxidation (M)
PPGVLKIFG	9	Q5U651	
QPESKRDIL	9	P12111	
HGSSAAAH	8	Q8NBK3	
SM(+15.99)ESFPPLAY	10	Q96M27	Oxidation (M)
LLDEPTNHL	9	Q9UG63	
ILDEPTNHL	9		
M(+15.99)EAIPVKQF	9	P23470	Oxidation (M)
ALFQRPLI	9	O60832	
KLSLVAAM(+15.99)L	9	P11021	Oxidation (M)
SVSGM(+15.99)TLKY	9	Q5VZL5	Oxidation (M)
AETPQAKKI	9	Q14978	
SPAPTHNSL	9	P49327	
NEVESGYHSL	10	O94900	
SLPHVGYLY	9	Q969F9	
NLAEKLIGV	9	Q86X10	
APGKKTSGL	9	Q9UBU8	
NEM(+15.99)M(+15.99)SKEQF	9	Q16236	Oxidation (M)
M(+15.99)PSLKHIVEL	10	O60518	Oxidation (M)
AFKEM(+15.99)FLDY	9	Q9Y2Y0	Oxidation (M)
ELPEDEEEK	9	P08238	
THRGPAKH	8	Q2M1K9	

TPGPAGSRL	9	P57086	
KPHPPPSGL	9	Q9Y5Z7	
VFLSIKGIYY	10	O00541	
SLDALITHV	9	P00325:P07327	
KILDLETQL	9	Q5BJF6	
SPFFREAL	8	Q9P1Z0	
RADEQAFLV	9	Q14865	
ALGTDAEGQKV	11	O00186	
TVM(+15.99)DSKIVQV	10	O15131:O60684:P52294	Oxidation (M)
SPFSSESRAL	10	Q99593	
LPSRPPSQL	9	Q96IT1	
AEGEIYTKL	9	Q8TAG9	
EENVGLHQT	9	P67936	
IEHGIITNW	9	P68133:P68032:P63267:P62736	
M(+15.99)M(+15.99)APLHIAV	9	O75762	Oxidation (M)
SADEVQRFF	9	P31943	
QPTVRITEL	9	P19387	
SIREAGGAF	9	Q9UII2	
SPARLSGL	8	Q14296	
RLFEPHEL	9	Q9Y3M8	
VLLDHLSLA	9	P29459	
QPYGNTRITF	10	Q9H094	
RVNDVPEEFLY	11	O95486	
STDYETHEF	9	Q16585	
IQHSEFEEV	9	Q9ULJ6	
KM(+15.99)LDEAVFQV	10	Q8NEY1	Oxidation (M)
KLAEIVKQV	9	Q8IW41	
KSLALLNY	9	Q9UBU8	
FLLDKKIGV	9	P78371	
AEITDKLGL	9	P61204:P84077	
VVKPPGSSL	9	Q92560	
M(+15.99)FFGGEGSLTY	11	Q9H8J5	Oxidation (M)
RAEKDRESL	9	Q14152	
AVDPNRAVPL	10	Q9Y3C1	
ALIEKLVEL	9	Q14181	
YLLPKDIKL	9	Q6T310	
GEIDGNKVTLDW	12	P19338	
EVLGLILRY	9	P78527	
ALHDILTEI	9	P40937	
LFETPILARY	10	Q08431	
KPAGPPGILAL	11	P35579	

KLIGDPNLEFV	11	P62826	
SLHPDLIYNV	10	Q9BR76	
SPYPGLRLI	9	P17405	
ALLQSDVRL	9	Q15149	
TVAEINATM(+15.99)SV	11	Q9NVA2	Oxidation (M)
ALSSLIHAL	9	P13010	
APSQGRPGL	9	P12109	
TERDKQSKW	9	Q8TEL6	
M(+15.99)LANDIARL	9	Q9H4M9	Oxidation (M)
KVM(+15.99)EIHSKY	9	P56524	Oxidation (M)
WLDKIDRRY	9	Q5VIR6	
SRGDFVVEY	9	Q9NQR1	
KIGDFGLATV	10	P10398:P04049:P15056	
KLGDFGLATV	10	Q8N568	
ALAGTFHYY	9	Q15738	
FAYDPSNYEY	10	Q9H6L4	
RLNKVIKSV	9	Q16531	
APALPPPAF	9	Q93062	
EEIFGTIGM(+15.99)	9	Q99873	Oxidation (M)
AEILQESRM(+15.99)	9	O75347	Oxidation (M)
NPRNFSDNQL	10	P37802	
RPFERRLYVL	10	Q12789	
IPQKQREITL	10	Q9NR30	
AAEESKVV	8	P29966	
NLLLDYEL	9	P12111	
LPFGKVTNL	9	P26599	
LPFGKVTNI	9		
AENPGKYNIL	10	Q00839	
SEGNRISGATF	11	P28074	
RLQESVM(+15.99)EA	9	O60216	Oxidation (M)
KLLDAYLLY	9	P61803	
IRKPYIWEY	9	P07814	
SASGPHSSL	9	Q9BRR6	
AEQDLIREL	9	Q9BXC9	
SADEPM(+15.99)TTF	9	P23193	Oxidation (M)
AAHEVGHAL	9	Q99542	
LLGPRLVLA	9	P49755	
NLAEDIM(+15.99)RL	9	P08670	Oxidation (M)
RVQEPGAVL	9	Q9NYB0	
RPAPPAQSL	9	O95208	
YLADIFTKL	9	Q49AG3	

KSTAISLFY	9	P13639	
LPRVPQNL	9	Q9NWF9	
LM(+15.99)YPYIFHV	9	Q9NZB2	Oxidation (M)
SLM(+15.99)SHAIEL	9	Q9Y263	Oxidation (M)
NEFEATQKL	9	Q16649	
YVM(+15.99)VGFFGY	9	Q9HBR0	Oxidation (M)
YLEPKLTQV	9	Q14997	
REFPDRLVGY	10	Q93063	
M(+15.99)NAKM(+15.99)HFRF	9	Q9HCL0	Oxidation (M)
HIENIVAV	9	O76021	
RPQKPSLEF	9	Q96AP0	
YYDPTTGLYY	11	P52756	
GILGIQPPSV	10	Q9UPN6	
SEKLDYITY	9	Q12874	
KPAASITSK	9	O43670	
KM(+15.99)TEVM(+15.99)M(+15.99)NT	9	Q53QV2	Oxidation (M)
KLDETGNSL	9	P11388	
NLDDGIDDER	10	P11940:Q9H361	
APVSPHSTF	9	O75159	
HPRSPNVLSV	10	Q16665	
RPGKIVQM(+15.99)	8	P62140	Oxidation (M)
LM(+15.99)HSIGLGY	9	P21359	Oxidation (M)
YYGYPPTY	9	Q8N755	
EEQTPQNKI	9	P00338	
TVM(+15.99)DKTRAL	9	Q5EG05:P29466	Oxidation (M)
LPAFKRRTL	9	Q00013:O14936	
NSDIPNHNM(+15.99)	9	Q9Y2H5	Oxidation (M)
APRVPVQAL	9	O75179	
AEGIPKLL	8	P12111	
SETGALKTF	9	Q96CW1	
RVIGTLEEV	9	P46060	
LPPLPKRPAL	10	Q9NR56	
VPGPRRSGA	9	Q99805	
EEIREAFRVF	10	P0DP24:P0DP23:P0DP25	
KPVYPGQTL	9	P51659	
KLADQIFAYY	10	Q96F07:Q7L576	
RPKTGFIRV	9	P07996	
SGDPKVIEV	9	Q9UPN3	
AAQKGRITQL	9	Q9H2K2:O95271	
M(+15.99)PLLRQEEL	9	Q9NZN4	Oxidation (M)
VLREIAEEY	9	Q00341	

N(+.98)YLHDFLKY	9	Q9UBU8	Deamidation (NQ)
RPSASTQAL	9	Q16531	
YLPYNHQHEY	10	O95864	
KM(+15.99)TAYITEL	9	P27540	Oxidation (M)
KLFEFM(+15.99)HET	9	O14980	Oxidation (M)
REAVTHIGRVY	11	P06730	
TELERKLTf	9	Q9ULG6	
RVSEPHRSI	9	Q68EM7	
EELSPTAKF	9	Q99612	
KRFGKAYNL	9	Q86XP3	
APGPRFLVT	9	Q6YHK3	
TLSEIKALL	9	P12814	
KLQSSEAEV	9	Q9P2E9	
LNLPYFLRY	9	Q5T4S7	
SPRVPLSAY	9	Q9UQ35	
KIALPHGQL	9	Q12796	
VVFNDPEKV	9	Q9H3M7	
HLLEPIYL	9	Q9BUB4	
NIYDSM(+15.99)PLRY	10	Q9UBQ6	Oxidation (M)
RISDVHFSV	9	O95503	
ALFEGKVQL	9	Q8NEZ3	
RPAVGHSGL	9	Q8IXZ2	
SLTSEIHFL	9	Q96SN8	
GLHGVTFGY	9	Q8NE71	
KM(+15.99)M(+15.99)DVTVTI	9	P69849:Q15155:Q5JPE7	Oxidation (M)
SLPKGVEEV	9	P50454	
KELGGADNIRKY	12	Q9NXU5	
ALDGKIYEL	9	Q13017	
SSLPTQM(+15.99)DY	9	Q9Y6A9	Oxidation (M)
M(+15.99)ISHIVTYY	9	P00395	Oxidation (M)
KLWEM(+15.99)DNM(+15.99)LI	10	Q96HT8:Q9Y605	Oxidation (M)
SAESKTIEF	9	P49792	
QPQHAFNAL	9	Q99967	
HPKNPEVGL	9	P07814	
M(+15.99)EPSTAARAW	10	P39656	Oxidation (M)
YRGQPGNAY	9	Q96PK6	
SEM(+15.99)TDKLGGL	9	P18085	Oxidation (M)
YRPPELLL	8	Q9NYV4:Q14004	
SIAEVVHQL	9	O14976	
LPSKPSSTL	9	Q9UNF0	
APQRAHEVL	9	Q9NVF7	

TEVDNYHFY	9	Q16236	
ASDQPSHSL	9	Q8TEW0	
RVDSHSKIL	9	Q13098	
HPNRTVPTAL	10	Q9BV87	
RPRPDDLEI	9	P17302	
TEDIQTRQY	9	P78362:Q9UPE1:Q96SB4	
YYSHLEGARF	10	O60287	
TDAEFRQM(+15.99)W	9	P53618	Oxidation (M)
RADAGHAGVSA	11	Q9Y657	
GM(+15.99)LEKAVFY	9	Q8NBN3:Q96K49	Oxidation (M)
M(+15.99)ALGLALRY	9	A5YKK6	Oxidation (M)
YLPEDFIRV	9	O43187	
HLDTDVKRV	9	Q9NPQ8	
SPRGFPLGL	9	Q9P270	
ALSDSIHTV	9	Q9UKN5	
LPPPPPHAL	9	Q9H1B7	
APDKKITL	8	Q12948	
KM(+15.99)AEM(+15.99)LVEL	9	Q96HT8:Q9Y605	Oxidation (M)
ALQEM(+15.99)VHQV	9	Q14511	Oxidation (M)
GEVSARQEL	9	P13639	
KPGVISVM(+15.99)	8	P12111	Oxidation (M)
RLGDGLGHVEL	11	P78357	
QM(+15.99)ISRIEYV	9	P48729:Q8N752	Oxidation (M)
SPQSPSSSF	10	Q9Y2J4	
GPSKVRVGV	9	P12111	
YVLPEVSKV	9	Q9UI10	
PEYHRKAV	8	Q9UBU8	
SEISSNKQNSY	12	Q8N3C7	
TPERHKPEL	9	Q7Z4H7	
RPEEAGRAL	9	P26022	
NADEARSL	8	O14602:P47813	
KTLESYLL	9	P23921	
REHDIAIKFF	10	P30260	
SPTIGKQL	8	P42285	
AEIQPQTQV	9	Q13308	
SADDGIRNL	9	Q13425	
VPNAGRGLVRL	11	P27695	
SARPAGDSV	9	P54198	
ITADVNHNL	9	Q16222	
ILDPHVLL	9	Q99567	
SLFEEM(+15.99)LQV	9	P61011	Oxidation (M)

REKDTVKKL	9	Q9P2E9	
SEIENVHGF	9	Q12959	
SPKPPSQQL	9	Q8IZD2	
KITDIHIKY	9	Q709C8	
IVDRPVTL	8	O96000	
SLLDKIIGA	9	Q6NZI2	
NKGPSYGL	8	Q15417:Q99439	
YTM(+15.99)KKVHAL	9	P10646	Oxidation (M)
ISRNSGSEL	9	P18850	
SPRKPTRKI	9	Q9UM11	
ALM(+15.99)GEHIHY	9	Q99967	Oxidation (M)
GVIPRSFEY	9	Q9NS87	
RADSSRASL	9	Q7Z7F7	
LPSENHKEM(+15.99)	9	Q15785	Oxidation (M)
RLAQHITYV	9	P33993	
GGPGTASRP	9	P08670	
RVKEKEAEL	9	Q9NVA2	
YLINEIDRI	9	Q03001	
KLKSTTESY	9	Q92802	
RIIDVVYNA	9	P62241	
KIIDEDGLLNL	11	P35251	
RPKDPNNLL	9	O00232	
KM(+15.99)IEDLQNEL	10	Q8IY18	Oxidation (M)
IIREPLTEL	9	Q9UPN3	
SEIGQKQKL	9	Q9HB11	
VLEDPVHAV	9	P61619	
RENDIKSYF	9	P39748	
LLQDKQFEL	9	Q96JC1	
KLGEIVTTI	9	P61204:P84077:P18085	
KLHGVNINV	9	Q9BWF3	
LLDKIKNL	8	P12111	
ALQGIHSI	9	Q9H6X2	
YYYDGDIGNYY	11	Q92769	
AVDGKEM(+15.99)VTM(+15.99)	10	Q96TA2	Oxidation (M)
SLLPPTALVGL	11	Q15436	
RM(+15.99)FDM(+15.99)GFEY	9	Q86XP3	Oxidation (M)
APFEPLASGIL	11	P01033	
QEELAEM(+15.99)KRY	10	Q9P0K7	Oxidation (M)
APVPIHPSL	9	Q8TAX0	
HFDSPYLLY	9	Q6P158	
SLSSFLHGV	9	Q96Q42	

TVGDKYM(+15.99)TV	9	Q02153	Oxidation (M)
SYHHGYPY	8	Q76M96	
AEADVASLNRRRI	12	P07951	
KPFDPARTPLVL	12	P56945	
SENDVIRLI	9	O94768	
TPKEKAQAL	9	P46013	
KTLEGEFSV	9	P25490:O15391	
GPSKKTAKL	9	Q12906	
AVSAVVHEY	9	P35610	
KPNSNQVNL	9	P48507	
SFLSIGARY	9	Q06203	
HEIHDPQHL	9	Q9H582	
RLLDLENSL	9	Q7Z7A4	
GPKPGGGPGL	10	P23246	
SPQRLRGLLL	10	Q96CG8	
TEN(+.98)KERKSF	9	Q9P0V3	Deamidation (NQ)
AEGGGAKFKKY	11	P49411	
AETGSGKTL	9	Q9GZR7	
APKAN(+.98)GTSAL	10	Q14978	Deamidation (NQ)
GLQHWVPEL	9	Q08431	
SPGQRAVLL	9	Q12772	
NEDEIKKAY	9	O75953	
TLM(+15.99)ERTVSL	9	Q8IWE2	Oxidation (M)
KLGPAPKTL	9	Q14527	
ALADGVQKV	9	O14791	
YPLESRPSL	9	Q14980	
KLFDRPQEL	9	Q5VZ89	
SEIAKEQGF	9	O75569	
ATSKVALVY	9	P06576	
SPREEPRAPW	10	P01889:P01893	
ITQGTPLKY	9	Q9Y618	
SVLEDPVHAV	10	P61619	
RIYQPKRYAL	10	Q52LJ0	
AQFEHTILL	9	P50579	
RVM(+15.99)APRALL	9	P10321	Oxidation (M)
YRAQPSVSL	9	Q96PK6	
SLLEKSLGL	9	O43324	
LPEWKENIM(+15.99)	9	Q9Y4C2	Oxidation (M)
SLSGELREV	9	Q8IX12	
KM(+15.99)DENQFVAV	10	Q16555	Oxidation (M)
SLYEENKLN	9	Q8IWJ2	

SFKDNSNLY	9	P17612	
ALGDRIKSL	9	Q9UKX5	
RPQIHGYDL	9	Q6IA86	
SPASKTTL	8	Q9HCJ3	
M(+15.99)PKVSGPDL	9	Q09666	Oxidation (M)
APRPPTAAP	9	Q8TF47	
AESAPAWRL	9	Q16342	
YRGPANGF	8	Q14444	
M(+15.99)PQKERESI	9	P38919	Oxidation (M)
QPAPTPLKM(+15.99)	9	P36956	Oxidation (M)
ALWEDEGVRA	10	Q5JWF2:P63092	
YFFGEGYTY	9	Q6P6C2	
RNADVFLKY	9	O75962	
SLHDAIM(+15.99)IV	9	Q99832	Oxidation (M)
VPENKGISL	9	P46013	
FLTEAALGDA	10	P48307	
TEIDARAGTF	10	Q13813	
HPEPPKAEL	9	Q13435	
NPDLRRNVL	9	Q15560	
SVQPDRELEL	9	O95071	
KPLGPTAAF	9	O60331	
GPKAGSRGL	9	P23560	
AASNVRGSL	9	Q16690	
RPDQLTGL	8	Q15149	
YYDGDVGNYYY	11	Q13547	
KLFPDTPLAL	10	Q12906	
KQHGVNVSV	9	Q14157	
NIRNIGKTL	9	P61247	
RAANFREGL	9	P35869	
KLLPSVTEL	9	Q8IZX4:P21675	
YVFDKRLRY	9	Q8N8U2	
AVRKPAISL	9	Q5VSL9	
SPAPDLAPR	9	Q504T8	
AEAKARAEI	9	Q9UNF1	
GVRPSPM(+15.99)QL	9	O00764	Oxidation (M)
KM(+15.99)DIDVPDV	9	Q09666	Oxidation (M)
RPRSALPRLLL	11	Q14332	
PPEYHRKAV	9	Q9UBU8	
GPM(+15.99)GPRGPP	9	P02452:P02458	Oxidation (M)
RPRPGTGL	8	Q9BVK6	
VPRPPQTSL	9	O95785	

QQFPEGLFY	9	P0CW20:P0CW19:P48059	
KILQELPSV	9	Q96ST2	
SPRIATTA	9	Q14980	
SPTAGKKL	8	Q8IWB9	
KPIQRTEL	8	P13639	
HGLPEQFLY	9	Q02880	
QVLGRFFLY	9	Q9BRB3	
IVDRGSTNL	9	P35611	
GLIDKVNEL	9	P61011	
AEAYGKKEW	9	Q5TFE4	
ALDSHTIV	9	Q02363	
SPKLPVSSL	9	Q99856	
YQLVKTYEY	9	O15397	
QPNPLPLRL	9	Q15154	
SPHGGALVM(+15.99)	9	Q8N201	Oxidation (M)
SPKPPTSM(+15.99)F	9	Q99081	Oxidation (M)
SELLSDVRF	9	Q9HOC5	
RTFPFSLQY	9	O75159	
LLAEKIYKI	9	Q09472	
RLQDELVTL	9	Q9UPN3	
KPLDGSAL	8	P12111	
KILEDVVGW	9	Q9ULW0	
ESKDPADETAD	12	P16949	
KLM(+15.99)DLDVEQL	10	P12004	Oxidation (M)
EELGNHKAF	9	P09543	
KEIEDRQKL	9	Q9NQZ2	
SPDSRNTLIL	10	Q13425	
TVSDAILRY	9	Q9Y5W9	
SEAAAPRAF	9	Q86X12	
NM(+15.99)M(+15.99)SGISSV	9	P68400	Oxidation (M)
TLVDEVFRI	9	P15927	
SETTAAHSL	9	Q9HC98	
GLIEILKKV	9	O14737	
APM(+15.99)SAAHQL	9	P54259	Oxidation (M)
KLLGELHTL	9	Q15154	
GPKRPGNTL	9	O75962	
KVREEFKEL	9	P02545	
TPAM(+15.99)QRKKL	9	Q9UG63	Oxidation (M)
EEM(+15.99)KKGSQL	9	P07814	Oxidation (M)
M(+15.99)IRAAPPPL	9	P10619	Oxidation (M)
RPQAPATVGL	10	Q15370	

KLSEYEM(+15.99)SI	9	Q8NBU5	Oxidation (M)
SISSITIRY	9	Q02763	
VPSESSEKAL	9	Q92541	
AESSGQKSF	9	Q5QJE6	
SLWGQPAEA	9	P29400	
GESVIGKTY	9	Q9P227:Q5T5U3	
RPPNYSGTIAL	11	Q13454	
APSGTRVVQVL	11	Q9UN70	
AFIEKHWTF	9	Q9H8E8	
QPTHPIRL	8	Q04917	
REEQKNLHFF	10	Q9UHY1	
RLHEKIQEA	9	O75683	
ALAAALAH	9	Q9NR30	
LLLLLLRSAGCW	12	Q9UI38	
RLIEVNGVDL	10	Q8TEW0	
AM(+15.99)NGHVPVAV	9	Q8IWZ3	Oxidation (M)
SPGVGKTSL	9	Q9NU22	
EEHGYSRQL	9	O75762	
SLSEKTVLL	9	P13987	
TPRPGSHTI	9	Q13131	
HPGPRPAL	8	Q9UGL1	
SPLKGLSL	8	P31350	
SPRVPGGSP	9	Q9Y6G9	
SPAAKSPSAQL	11	Q14203	
TEVTGHRW	8	P35613	
LLIENVASL	9	P07203	
AEAPSQVRL	9	Q9P2E5	
RLHERLVAI	9	Q15149	
SPNSYFM(+15.99)DV	9	Q71UM5:P42677	Oxidation (M)
FQDPVPLTV	9	O15164	
RLQEEINEV	9	Q70UQ0	
RQLEEEGITFV	11	Q92503	
RVAPPRALL	9	Q9BXF6	
KPSPIYGGL	9	P03923	
AGDHPVEL	8	P09486	
VAETPDIKL	9	P46782	
SADEVQRF	8	P31943	
APAPSKGQTL	10	Q6ZRI6	
AEIDSIHQL	9	Q9P2W9	
SPSSLIGQAL	10	Q96AY4	
NEIDDM(+15.99)QVF	9	Q86T65	Oxidation (M)

SFTTFTVTKY	10	Q03135	
SELQSIQRL	9	O15119	
YM(+15.99)SGYHTGY	9	Q16637	Oxidation (M)
TLLEDGTFKV	10	Q9HBL8	
RLLKSKLYL	9	P10620	
TPVAGSQSL	9	Q9UQ35	
SPQKGRGGLG	10	O43680	
SLDTQPKKV	9	P15923	
SRPGGYGYG	9	P98179	
KPASFM(+15.99)TSI	9	P53396	Oxidation (M)
APRPTIINA	9	Q5JSZ5	
ALLGDLTKA	9	Q92973:O14787	
IVRKIGTNL	9	Q15390	
VPNKGYSSL	9	Q8IW50	
TPSRKGLVL	9	Q9H4L4	
FVSEIVDTV	9	P12111	
TSDNPVVQL	9	O00505	
KENSYVKKF	9	P28289	
SLYGGNAVVEL	11	O75787	
NSLPGNFFY	9	Q15404	
VLLDYHLNY	9	Q06787	
YLDPAQRGV	9	Q96S19	
QLVDLHARY	9	P36969	
KALEQILEY	9	P51946	
RPAAGFVAGL	10	O15427	
M(+15.99)PRGPTLGHVSL	12	Q96TA2	Oxidation (M)
SLEHLPANLFY	11	O60346	
RVFPYSVFY	9	O15118	
ALLESLRQA	10	P24385	
RPENRSFYL	9	Q9BQ70	
EELEGIQKL	9	Q6AZY7	
SLINVGLISV	10	Q92688	
TIDGPSKV	8	P21333:Q14315	
KVFDQYLNF	9	Q8WVM8	
DEQDSVRL	9	P30153	
SPDAARIRV	9	P35318	
SLENGAHAY	9	P26358	
TPRLPQRNL	9	O43516	
SPRSITSTP	9	Q9P0K7	
IPDASKLQFIL	11	O75935	
GEVGKTTGI	9	P53396	

RPAEKLSTA	9	O95563	
AEIENIKEL	9	Q86VG3	
AM(+15.99)M(+15.99)AKAEYL	9	Q15008	Oxidation (M)
KLGDIM(+15.99)GV	8	Q92620	Oxidation (M)
YFPHFDLSH	9	P69905	
RLVEIQYEL	9	Q14161	
KM(+15.99)FSDEILL	9	Q724H8	Oxidation (M)
TEQSQIKGY	9	Q9NZ42	
FLNKEITSV	9	Q9UI10	
RPANHRQLTF	10	Q9UNM6	
KIM(+15.99)EQSPDM(+15.99)	9	P35716	Oxidation (M)
QLFEKQLQL	9	Q05209	
KPGM(+15.99)KLTTEL	9	P12270	Oxidation (M)
SETGRSKGY	9	Q14498	
QFTAM(+15.99)RDLY	9	P62834:P61224:A6NIZ1	Oxidation (M)
SLIDKHLNL	9	A1A4S6	
FPNKPPLTF	9	Q86U86	
LDKKVEKV	8	P08238	
KIIDIQEKV	9	P12270	
NLFGGEPLSY	10	P02786	
HLDEAQRLLY	10	Q8IVP9:Q8NEK5	
RVRVSGQGL	9	P21333	
TLAETLVNL	9	Q8NB46	
RVAPEEHPVL	10	P60709:P63261	
NEITGHTY	8	Q6ZUT9:Q6IQ26	
RWNPLPLSY	9	Q9UPW6	
RPPVGHFSF	9	Q8NEZ4	
KM(+15.99)QETLVGL	9	Q9Y6D9	Oxidation (M)
GEVEGHPVW	9	Q8N1F7	
RADDARQL	8	P63096:P08754:P04899	
NENDIITHF	9	Q9ULK4	
SVKSPNQSL	9	Q96B01	
FLDEPTNHL	9	Q9UG63	
GPPGTGKTAL	10	Q9Y265	
GLKEGIPAL	9	P13639	
KIDEKTAEL	9	Q9BZD4	
VPNAGRGLV	9	P27695	
SPSRLPATSL	10	Q8TEJ3	
KEFLKPEKF	9	Q8N543	
SVAKESIPL	9	A6NHX0	
KVLPQTILY	9	Q9NWF9	

KVIDYVPGI	9	Q7Z3E1	
AFDIPRFVY	9	P56282	
AEALVSKGL	9	Q7KZF4	
SPDTKKIAL	9	Q86TI0	
TIM(+15.99)EDYFAY	9	P51532	Oxidation (M)
AM(+15.99)IELVERL	9	Q96DX7	Oxidation (M)
QLQEKLVAL	9	Q86UU1	
VPESHRKPL	9	Q01850	
SPQDKRTL	9	Q8TC07	
YLDSM(+15.99)YIM(+15.99)L	9	Q13443	Oxidation (M)
EEM(+15.99)RKIIGRY	10	Q9UG63	Oxidation (M)
RLDDAIHVL	9	Q99081	
NPKRQTLVF	9	Q9GZR7	
SFFDLNPKY	9	Q9BQL6:Q96AC1	
RPDPNTGKVH	10	Q15393	
RLFNDPVAM(+15.99)V	10	Q92685	Oxidation (M)
RPKGESKDL	9	Q16531	
ALDPIIHGL	9	P23470	
IPQKFHRSV	9	Q00341	
RLQEELEKL	9	Q9P2E9	
SLPSSGRSSL	10	O95817	
SEYDRGVNTF	10	P28066	
ALSEQILM(+15.99)V	9	P48552	Oxidation (M)
ALASLIRSV	9	Q9Y646	
RQIDQFLVV	9	Q9BTC8	
SRPPLLGF	8	Q15386:Q7Z3V4	
RLLEIALGF	9	P35555	
M(+15.99)FTELAILY	9	Q00610	Oxidation (M)
HYNNIM(+15.99)ALY	9	Q16531	Oxidation (M)
SPNGKESDL	9	Q99575	
KLRGVSLAA	9	P29400	
KPAGPPGI	8	P35579	
SVLM(+15.99)KALEY	9	Q14008	Oxidation (M)
AM(+15.99)TQLLAGV	9	P24821	Oxidation (M)
RPM(+15.99)GAGEAL	9	Q12906	Oxidation (M)
SPDGVKRITA	10	Q8TAT6	
SIHDVTFQV	9	Q9P2D0	
RLAHYIDRV	9	Q03252	
RPQTGLSFL	9	P43686	
IPYGERITL	9	P49750	
TEIDKPSQM(+15.99)	9	P02751	Oxidation (M)

GESGGSRVF	9	P48681	
HILGAAYKY	9	Q92922	
SIM(+15.99)DEFQEY	9	Q8WUH1	Oxidation (M)
HLSIINEYL	9	P46940	
LARPAALTAL	10	Q8NBS9	
YVDDTQFVRF	10	P04439:P01889:P01893:P10321	
EEETTRISY	9	O43396	
RETSM(+15.99)VHEL	9	P61619	Oxidation (M)
GTLRETLEY	9	Q9H4M7	
VIDEPVRL	8	P46013	
RLWPKIQGL	9	Q9NX09	
RPYSQVNVL	9	P46939	
NPKNVNRVL	9	Q14690	
RPRLTSNEKL	10	Q9Y296	
KPKTPSLTVF	10	P24928	
VLISGVVHEI	10	O43462	
KLTDVGIATL	10	Q86TB3	
RPAGAQQHHQL	10	Q6SPF0	
TLDAGNIKL	9	O75400	
KLTDYQVTL	9	Q9H1H9	
AM(+15.99)NPIIYSY	9	Q92633	Oxidation (M)
RPPLGSTEL	9	P42166	
SESDLNHSF	9	Q9Y385	
SEPDSTHVY	9	Q05707	
VM(+15.99)LDVPIRL	9	Q9UJF2	Oxidation (M)
FM(+15.99)LPDPQNI	9	Q8WW01	Oxidation (M)
KVYENYPTY	9	P35659	
KPM(+15.99)TTKDLL	9	P35269	Oxidation (M)
LPLRSSSSL	9	O95136	
SPNQARAQA	9	P26599	
IM(+15.99)AQLPQEQKA	11	P35221	Oxidation (M)
M(+15.99)PSQFRDIL	9	Q09472	Oxidation (M)
QPRRPRNLAA	10	Q00325	
IVDRPVTLV	9	O96000	
RIAHRIQEL	9	P51531:P51532	
RPNPGFSM(+15.99)L	9	Q9HDC9	Oxidation (M)
SM(+15.99)VDVVM(+15.99)LL	9	P46379	Oxidation (M)
RLAELEEFI	9	Q00610	
NIFPNPEATFV	11	Q9Y5B9	
LLIDDKGTIKL	11	P06493	
M(+15.99)VQNNRIEL	9	O75762	Oxidation (M)

RLQEKVESA	9	Q09161	
EKSIDLIQKW	11	P49585	
VPAERREKL	9	Q6NZ12	
FLLEKGYEV	9	O60547	
VVLQGANLYGY	11	Q9NYZ1	
SLVAVELEKV	10	Q9H9E3	
AEIQKDVQTY	10	Q9HBK9	
SPSSHPAL	8	Q96TA1	
YPDTHRHL	9	P04040	
ALAEALKEV	9	Q6PI26	
AEDGLKHEY	9	Q9UQ88:P21127	
EEAQIIQRRL	10	Q9NQZ2	
KETEIVKKL	9	Q03701	
SLFGKKYIL	9	Q14192	
VIAEILRGV	9	O00567	
IIDKEVSL	8	Q9NUQ6	
RLYGYFHDA	9	O14965	
GLTEALHFV	9	P07199	
SPRGSSTSL	11	Q9H7P9	
EEFSRAAEKLY	11	P49458	
RVRDINEAF	9	Q99081:P15923	
REAELYHTL	10	Q7Z460	
LLDYPPDRVTL	12	O60568	
RPSGEAFVEL	10	P31943	
KAKPVTTNL	9	Q8N0T1	
LLHEEHSRAAA	11	P02462	
RPNDIM(+15.99)AEV	9	Q13131	Oxidation (M)
APNYRLKSL	9	Q16594:Q9HBM6	
HPM(+15.99)YGNIL	8	P55735	Oxidation (M)
RPGYRVKEI	9	O00422	
YPDDRTARTL	10	Q5VWQ8	
SYNDYVREF	9	Q96MU7	
KLADIQIEQL	10	O60925	
RPNLSHTGRGIL	12	Q13356	
QEFEEKSGRTF	11	Q9NP77	
SPGKGVVSL	9	Q14186	
KIAQFLESV	9	Q9NU22	
LVYPGDPLRF	10	Q9BSV6	
QQALQKKL	8	P07951	
SPRPPLISV	9	Q9H0R1	
SPRSSDLVKL	10	Q9NVH2	

SLAEKIQAL	9	Q9NWX5	
QEAASLLGKKY	11	P49321	
RPAQLHIGF	9	P22314	
RTAGPHTQF	9	O75448	
SEHGAQHIIY	10	P28347	
GLQKEIAEL	9	P15408	
RPKVPVLQL	9	P56192	
KLIPFLEKL	9	Q05D32	
NYIEFTRIL	9	P19105:O14950	
KPLSNM(+15.99)KIL	9	P09874	Oxidation (M)
FRLDTPLYF	9	Q32P28	
PDGSKPVKV	9	Q01995	
QESKGISQF	9	Q14865	
TLVEAIKQV	9	Q9HCE1	
YRGIFPVL	8	P14618	
SPRNLFETSFAF	12	Q9UBF8	
RLYDGLFKV	9	Q16531	
SPTRNSKL	8	Q9Y5W7	
SPGERGEV	8	P08123	
SEQFLIGL	9	Q8N1G2	
RETDVAKKI	9	O14802	
ALVSSLHLL	9	Q9Y678	
SQDDVSRF	8	P12111	
SIM(+15.99)EKTQSL	9	Q05707	Oxidation (M)
TPSDPRRRL	9	O60716	
GEVSGLTKDF	10	Q5VV42	
SARPPDSGL	9	Q9C0G0	
YYSGLIYTY	9	P35579	
SPPLRPRLLL	10	O75882	
NQM(+15.99)PTQPQV	9	Q09472	Oxidation (M)
KPAATKPAT	9	P12111	
APSEKFLNM(+15.99)	9	Q9H3K2	Oxidation (M)
SPTKKVAVA	9	P19338	
SLFGGSVKL	9	Q8WUM4	
KPATKTAEL	9	P18583	
M(+15.99)M(+15.99)LDDLLQL	9	Q99832	Oxidation (M)
AYRQGHFTY	9	Q96SB8	
TPNKSGKTTL	10	P49790	
APNLKILNL	9	Q9UBU9	
RPFLERRDL	9	Q9H2G4	
KLPEKWESV	9	O76021	

KPQSKELTL	9	Q9H4H8	
YIDEGRLVL	9	P35610	
LILDPIFKV	9	P13639	
APEKKPLV	8	P35221	
LPNRHIGISF	10	Q14315	
KENM(+15.99)AGKGF	9	P40763	Oxidation (M)
SLDAKEIYL	9	Q8IXB1	
YLPPATQVV	9	P38919	
TLAQRVKEV	9	Q9Y679	
EEFTHKDGVV	10	Q6P2Q9	
KTM(+15.99)EDTLM(+15.99)TV	10	P53365	Oxidation (M)
VM(+15.99)LGTQLLY	9	Q9UBU8:Q15014	Oxidation (M)
SPHKNSISL	9	Q9NP74	
GPAPGRLGP	9	P78324	
AM(+15.99)DSPLLKY	9	P54920	Oxidation (M)
SPVFKTVL	8	Q9HAV4	
QELDSTDGAKVF	12	P61011	
RLLDLENSLL	10	Q7Z7A4	
RPSHPLKSF	9	Q92859	
QPRLM(+15.99)EPIYL	10	P13639	Oxidation (M)
RLLDYVVNI	9	Q63ZY3	
VM(+15.99)FGGKQVVV	10	O43865	Oxidation (M)
KLAESVEKA	9	Q9Y5B9	
EEQRFNNFL	9	P13010	
APFQGDQRSL	10	Q00978	
KLADDVDLEQV	11	P55072	
GTFTGEALQY	10	P12109	
GLTDIRTLV	9	Q15751	
FTSVSRPGR	9	P04439	
KVANIILSY	9	Q96JI7	
SESSHKGFHY	10	Q8IVM0	
NIIPRFVQV	9	P0DMP2:O75044:PODJJO	
SKERSGVSL	9	P16402	
SPDLKM(+15.99)GNL	9	Q15596	Oxidation (M)
SNLDLQLEY	9	Q8IVH8	
AEIEIVKDL	9	P13674	
SEPESIRKW	9	P09496	
HPHGSYSRL	9	P40306	
LLIDDEYKV	9	Q8N766	
TADGKTYYY	9	O14776	
EEVQFLKTL	9	P49588	

SM(+15.99)FGKLIDL	9	Q96SY0	Oxidation (M)
HRNEVTVEL	9	O00505	
LVDDNYFYL	9	Q6P2Q9	
NEVRKVKKY	9	P12268	
RVKEIGSTM(+15.99)	9	O00422	Oxidation (M)
EDFESDYKKY	10	Q8NFA0	
ALLPVVHEY	9	Q9UNH6	
DSSEKFLR	9	P02808	
SLARAGFY	9	Q13489	
RPYEKGRLL	9	P31689	
ANPRSVSSL	9	Q8NEY1	
APPSKPTVL	9	Q15554	
AQQPHIAGV	9	Q8TB72	
NSDRPGVL	8	P57737	
VARVPGSQL	9	O75151	
APRGPGAEL	9	Q96FS4	
SLLDLEQKL	9	P56589	
IPRAALLPLL	10	Q92743	
ARADSYEY	9	Q9UKM7	
M(+15.99)LIEVIEKL	9	O00203	Oxidation (M)
RVREM(+15.99)FSQL	9	Q9UPN3	Oxidation (M)
NPNM(+15.99)RLSQL	9	Q96CT7	Oxidation (M)
KLDEILKEI	9	Q9ULV4	
SM(+15.99)AQAM(+15.99)KGV	9	O43633	Oxidation (M)
QPRSPVPSAF	10	Q9UQ35	
ARPATSTL	8	Q04637	
TPSSKPTLL	9	P27816	
VLIPKLPQL	9	Q8N138	
ALLGKIEKV	9	Q10471	
LLSDTVQHL	9	P42226	
LSADVFARY	9	P56192	
RVREIINAV	10	P31350	
HRPPVIVL	8	Q92743	
GLKNGVPAV	9	P53621	
ALASVIKEL	9	Q8WYA0	
IPRAALLPL	9	Q92743	
GLGPPGRSV	9	Q68D10	
RLFDQAFGL	9	P04792	
RLYTEPLL	9	Q969N2	
APDQEVRI	9	P36405	
ASDDGTVRI	9	Q9H7D7	

KM(+15.99)DDDVFL	8	P24928	Oxidation (M)
SYPIKLFY	8	Q9H9S3	
AEQEDRALRSF	11	Q00341	
AEM(+15.99)ETQKEKI	10	Q9NTJ3	Oxidation (M)
GENPASHSY	9	P13995	
SLIERDLKL	9	O60287	
EGPEGLKLI	9	Q8N335	
ILNDM(+15.99)VWKY	9	Q9ULK4	Oxidation (M)
SLFERLVKV	9	Q9P2E3	
SPRQALDF	9	Q8WTW3	
HPSLPASSL	9	Q53ET0	
RLALFNPDV	9	O00483	
SPHHALGL	8	Q96F45	
TPRSKGQESF	10	P06748	
EEM(+15.99)LEIRQL	9	Q9BZE4	Oxidation (M)
RLFADILNDV	10	Q96GQ5	
SVLPGELM(+15.99)RGY	11	P35251	Oxidation (M)
ALLEQVNAL	9	O75935	
SPRKPLFL	9	Q8IZC4	
LLAEKVEQL	9	Q13454	
RTDGKVFQF	9	P83731	
KLQDQIFNL	9	O43665	
VEELFERKY	9	P78347	
RAEELGQEL	9	Q14980	
NPRDPPSWSVL	11	Q9Y490	
GLFQGKTPL	9	Q9ULW0	
KLISELQKL	9	Q5UIP0	
ASRLPPATL	9	Q9NU23	
RPETGRPRP	9	P46783	
KIQEILTQV	9	O00425	
SLFSEETPVVL	11	O75787	
GM(+15.99)LEKIDM(+15.99)I	9	Q9NTZ6	Oxidation (M)
AM(+15.99)GIAPPKV	9	O43395	Oxidation (M)
RLIQESPTL	9	Q6P1A2	
RLLEGVNQL	9	Q10570	
RPSDANRKEM(+15.99)L	11	O43847	Oxidation (M)
SM(+15.99)ASLVIFM(+15.99)	9	Q9UKV5	Oxidation (M)
SPAARTGV	8	Q09666	
RIQHPSNVL	9	P14866	
VADPDNPLVL	10	P39656	
NVIERLQEA	9	Q9UBK9	

SLADDSVLRL	11	Q9NU22	
FVRKATREM(+15.99)	9	O76074	Oxidation (M)
ALAPGVRAV	9	Q9Y508	
FLLPHPGLQV	10	Q6DD88	
RM(+15.99)DENSPLM(+15.99)Y	10	Q14643	Oxidation (M)
RPAQGVVTTL	10	P02751	
EENM(+15.99)RGVRF	9	P13639	Oxidation (M)
NILEQIVSV	9	P48058	
SLYEM(+15.99)VSRV	9	Q08945	Oxidation (M)
TVRSTGSATSL	11	Q55W79	
SRPPHLVF	8	P15848	
NDN(+.98)YKLIELI	10	Q7Z333	Deamidation (NQ)
GFASFSDYYY	10	Q96BY9	
SENAGIQSL	9	Q9P0K7	
NERDGVITYSY	10	Q9NRX5	
KLDENSAFL	9	Q13813	
SLEENLEKI	9	O60812:PODMR1:B7ZW38:B2RXH8:P07910	
TPAEVRQVL	9	Q9H9E3	
QLSPRLLEV	9	Q460N5	
TLIENLTEI	9	P16234	
RPPLEEKAL	10	Q9HB19	
KILDIGLAY	9	P49754	
M(+15.99)TRGNQREL	9	P84101	Oxidation (M)
LPRGM(+15.99)QPTEF	10	P28062	Oxidation (M)
RVDEEPTTL	9	Q9Y4E1:Q641Q2	
IPKSAVGTL	9	P05556	
KRPPTM(+15.99)TV	8	Q8WW11	Oxidation (M)
ALNENINQV	9	Q9BZ23	
LLM(+15.99)EHIQEI	9	Q96SN8	Oxidation (M)
RPSAAGINL	9	Q08211	
TIRETTEEM(+15.99)	9	Q9NTK1	Oxidation (M)
SLEENIAKA	9	Q93074	
SPQKSTVL	8	Q9NVP1	
SEAVYQHLF	9	O43299	
GPRTAALGLL	10	Q14257	
SPAKPKAV	8	P16401	
STSRSLALM(+15.99)Y	10	O95721	Oxidation (M)
SLADVHIEV	9	O14981	
SEAGPNQTF	9	Q16666	
NENDIRVM(+15.99)F	9	O95319	Oxidation (M)
KPSSFTTSF	9	Q8NI77	

VPNQKRLTL	9	O60488	
KYM(+15.99)DVQFDF	9	O00159	Oxidation (M)
VLQGKLAEV	9	Q9UPN3	
KPEEQRQAL	9	Q15149	
ASPKSPTAAL	10	Q53EZ4	
GPHPGTSDL	9	Q86XN7	
SLLDDLHSA	9	Q9C0D3	
LPTEKEVAL	9	P27816	
GGPGTASRPSS	11	P08670	
IPEEQRAQL	9	Q15149	
KLAPSEYRY	9	Q9BS26	
AENPDM(+15.99)LKKL	10	Q7L311	Oxidation (M)
RLPDPFAKV	9	Q9HAU4	
AEM(+15.99)QWKRDY	9	O96000	Oxidation (M)
KLLNVDPDNV	10	P24522	
AENGLTVRL	9	Q12789	
APTPRIKAEL	10	Q04727	
TEYYDGKTRRY	11	Q8N3C0	
SPSAAQKAV	9	Q15434	
SYQGKGFVM(+15.99)	9	Q92692	Oxidation (M)
LPDDVRRRL	9	P26358	
HPARPTVESF	10	Q15149	
LPSASHFSQL	10	Q9Y520	
SEKLGVM(+15.99)TY	9	Q96TA2	Oxidation (M)
FYEHAQTYT	9	Q7Z392	
ILAQEIVKA	9	Q9UBE0	
AEQQEKM(+15.99)LI	9	O76074	Oxidation (M)
QEPDFEEKRF	10	Q86VM9	
RLLETVLGY	9	Q9H583	
RPVTPITNF	9	Q7RTN6	
VPAGLPRSL	9	Q7Z5L7	
AAM(+15.99)AGRLM(+15.99)I	9	Q9NRN7	Oxidation (M)
SENSVRDIF	9	P61587	
VADGYPVRL	9	Q8WX93	
KAM(+15.99)GIM(+15.99)NSF	9	Q99879:Q93079:Q99877: O60814:P62807:P58876:Q 5QNW6	Oxidation (M)
QPLLREVL	8	P28070	
KLRDEIM(+15.99)AL	9	Q03001	Oxidation (M)
AARAIISEL	9	P78344	
VTM(+15.99)DVHSRY	9	Q9H0A0	Oxidation (M)
SLAGDVALQQL	11	Q15021	

RPAGGNGSGL	10	Q5JSZ5	
TPSSKPTL	8	P27816	
SPQRLRGL	8	Q96CG8	
LPNQVHRKSV	10	Q15019:Q99719:O43236	
APRHGSLGF	9	P39023	
AM(+15.99)FGKLM(+15.99)TI	9	Q29RF7	Oxidation (M)
AEM(+15.99)EVKRESY	10	Q92575	Oxidation (M)
SPSSIRSL	8	P33992	
RARHTQAEL	9	P55268	
HPENNRPII	9	Q7Z2D5	
SLLQTLYKV	9	P46060	
KIM(+15.99)DATNILV	10	P40763	Oxidation (M)
AVRPASLNL	9	Q96N67	
KPRPPPLSP	9	Q15642	
QLTNVILHL	9	Q9NQG6	
ALSEELVQL	9	P33176	
PGPRRSGA	8	Q99805	
RTIAKRLVL	9	Q9H633	
NEAFGDTKF	9	Q13263	
GPARAGSVV	9	Q71SY5	
TLSERLWGL	9	Q9NS69	
KPEQAQKAL	9	Q9Y4C8	
NLASFIEQV	9	P61457	
RPSSGDRNL	9	Q9NS73	
KQYGNEVFL	9	P50990	
SRVDDLYTL	9	Q9Y3M8	
SFHDLKRY	9	Q2LD37	
SIDRNPPQL	9	P52630	
GIDRPAEL	8	Q8IWZ8	
RPSQLGQVF	9	Q93073	
KVAPAPAVV	9	P62424	
KLADRSKIYM(+15.99)A	11	O00391	Oxidation (M)
RPFLRRL	8	P51531:P51532	
RPFLRRI	8		
AEEVQKKKRTF	11	P62841	
GPRGPPGAA	9	P08123	
RPPPPNSVI	9	Q14186	
RPAPLSAEL	9	P53602	
YRPVRQYVY	9	Q9NZB2	
LGDFLQLHI	9	Q9Y4C0	
KIGEGTYGV	9	P06493	

VLGEFLKEI	9	Q08AM6	
KLIDKLDL	9	Q8N110	
PDAKIRIF	8	P27635	
KVRPSTGNSA	10	Q9UJX3	
APHGFAEEL	9	P51805	
EEM(+15.99)KSGLKDF	10	P49792	Oxidation (M)
SRAPSTYTY	9	Q86UC2	
KVKKPATAA	9	P16402	
SEVGTKKEL	9	Q9BRH9	
PDLLPGRKY	9	P02751	
VPDTSRIYVAL	11	Q9UBK9	
M(+15.99)VREPG EGL	9	Q15113	Oxidation (M)
RLNEAAVTV	9	O95273	
VAGWPARA	8	Q5VWQ8	
RRFFPYVY	9	P20618	
SPALKRLDL	9	Q8WTW3	
APQVHIFSL	9	P30559	
TLDP RSFLL	9	P25116	
APGNGQRSL	9	Q86UU0	
SPTGGRM(+15.99)SVF	10	O95486	Oxidation (M)
YPNKPRSGGM(+15.99)L	11	O75762	Oxidation (M)
ALNNLLHSL	9	Q96Q15	
ILFGHENRV	9	O14775	
SLLESVQKL	9	Q13535	
KLYGSIFTGA	10	Q63HK5	
PTRAGRGH	8	Q92618	
SELEKTFNL	9	P23497:Q9H930	
RPKEDTQKL	9	Q8NFG4	
APASRLAL	9	Q96S16	
NPRIGTFDL	9	Q14739	
SLIKQIPRI	9	P62906	
SLFPGKLEVV	10	Q13045	
ASDGTVRL	8	P31942	
SEIKPPNLF	9	P49902	
QIRNPM(+15.99)KAM(+15.99)	9	Q9NRW3	Oxidation (M)
RPDLREKVI	9	Q15008	
NIDGRRKIAF	10	P62269	
RPKR PANQF	9	P67870	
KHNLKRETL	9	Q86UW6	
HEM(+15.99)EKAF AF	9	Q9H000	Oxidation (M)
APDSQRLKL	9	Q9UQE7	

ILSDVHNRL	9	Q8WUJ3	
SLQDLIEKV	9	O94979	
KVLNELYTV	9	Q7Z6B7	
RPRKAFLLLL	10	P13667	
HPRTGGERGFL	11	Q16394	
RIDFSKLTSL	10	P26599	
RVEEGVPQV	9	P12111	
KPSVKGRAQL	10	P40227	
ALPDQGIACA	10	P78318	
YPRKGRIAV	9	Q16555	
MPKKDARTL	9	P29083	
RTIATALEY	9	P78537	
IPRGPYAEV	9	Q8NEY1	
VLLGKVYVV	9	Q6TFL4	
TINGHNAEV	9	P22626	
REIEGRAI	8	P19338	
RVTQVRKEL	9	Q9NQG5	
GVGPV PARA	9	Q8TB96	
HGGSSM(+15.99)IM(+15.99)	8	Q13224	Oxidation (M)
VLQDIQVM(+15.99)L	9	P33992	Oxidation (M)
HLLSDELLL	9	Q9UKT7	
QLDDLKVEL	9	P42766	
SELPTAKEY	9	Q93009	
RLIDLPGTEL	10	Q14CB8	
FLWKEPATM(+15.99)	9	Q9BQ70	Oxidation (M)
RAARGAPPL	9	Q8TF74	
VADPPIARL	9	Q14697	
ILDAGGHNV	9	Q99460	
IPYHSEVPVSL	11	Q01082	
AEKREHEKEVL	11	P16949	
SPSSHPSTSF	10	Q8WYQ9	
RLNKVISEL	9	P05387	
SPLGKGNSSEL	10	Q8IVE3	
RPTLWAAAL	9	P17936	
RVIGDLDTV	9	Q96AX9	
YLHNQGIGV	9	O60566	
LPQGQGSQ L	9	Q9C0C2	
RPKAGDTREV	10	O15061	
YVRDFVM(+15.99)NL	9	P12111	Oxidation (M)
NPNHRTPTL	9	P29590	
RLYSVSYLL	9	P21333	

M(+15.99)PSSAVRVL	9	Q8N0X2	Oxidation (M)
KLIANNTTV	9	O96019	
GVL DGRFHY	9	Q9NW08	
RAAPFSLEY	9	Q15181	
ALVVQVAEA	9	P07686	
QLLETNLQL	9	Q14203	
GPRRLLLVAA	10	P25116	
GPLDKWRAL	9	P27635	
RLSEFIPAVF	10	Q12768	
KSLHGHPR	8	Q7Z4N2	
HLLPEAWAY	9	Q96H72	
APRTNGVAM(+15.99)	9	Q92567	Oxidation (M)
VPVAKTAEI	9	P49790	
TLSDIKALI	9	O43707	
TVM(+15.99)ETVTTV	9	P11532	Oxidation (M)
APVGGHIL	8	O00159	
SILDIVTKV	9	Q52LD8	
APPPPLPKA	9	O95359	
QGVVTTLENV	10	P02751	
KM(+15.99)ADIEYRL	9	P40937	Oxidation (M)
TENTEENRRF	10	P04075	
HLLNESPM(+15.99)L	9	Q92621	Oxidation (M)
QVFEGHthy	9	P35606	
AVKTFNSELY	10	Q9UPN6	
IPELHRPVV	9	Q9H8H2	
EEEEESHKKF	9	Q9BQI5	
RVAEFHTEL	9	P48556	
GLLTEIRAV	9	Q9UEE5	
SPRGEASSL	9	Q8IY57	
SDPFTHLAP	9	Q9UPY3	
APRPPSSSP	9	Q5H9R7	

**Tableau 2** - Immunopeptidome des cellules MRC5 infectées (protéome humain)

Peptide	Longueur	Source	PTM
KLLPALLQHV	11	P14210	
YFYDPDVGNFHY	12	O15379	
HYFSDPFLKY	10	Q15165	
M(+15.99)M(+15.99)NGGHYTY	9	Q99497	Oxidation (M)
FYEVFKVLY	9	Q00325	

YYINKISSTLY	11	Q8N6L1	
GGPGTASRPSSS	12	P08670	
YYDGDVGNYYY	12	Q13547	
YYDGDIGNYYY	12	Q92769	
KYFDEHYEY	9	P33552	
APRPASGPIRP	11	P26373	
YYRYPTGESY	10	Q16875	
LLLLLIGVCWC	12	Q86SU0	
GYYGKGLQY	9	Q4LDE5	
HVIDVKFLY	9	Q16531	
EEVHDLERKY	10	P55209:Q99733	
SFAARSFYY	9	P49327	
LPRPPPEM(+15.99)	9	Q00325	Oxidation (M)
YWGEIISQQY	10	P14735	
AM(+15.99)SSKFFLV	9	P41221	Oxidation (M)
AASKERSGVSL	11	P10412:P16403:P16402	
VLLDKIKNLQV	11	P12111	
NYLHDFLKY	9	Q9UBU8	
FYLSTPENYHY	11	Q9HD67	
TSFDKFKYK	9	P45985	
ALFPHPHIHEY	11	Q86V87	
EEQQALQKKL	10	P07951	
KVLDFEHFL	9	P60660	
RPRPGTGLGRVM(+15.99)	12	Q9BVK6	Oxidation (M)
GPREFRQL	9	P57086	
SIGDIFLKY	9	Q9H8V3	
SVSDQFYRY	9	P55010	
AEDKENYKKF	10	P08238:P07900	
SLFPHPHIHEY	11	Q5W0V3	
VM(+15.99)DSKIVQV	9	O60684:O15131:P52294	Oxidation (M)
TEANVVRKF	9	O00391	
FGGPGTASRPS	11	P08670	
M(+15.99)TNPHNHLY	9	Q15006	Oxidation (M)
M(+15.99)VTEIRLKY	9	Q15393	Oxidation (M)
LPYNHQHEY	9	O95864	
AFFAERLYY	9	P20073	
APSRNGM(+15.99)VL	9	P26373	Oxidation (M)
HIKVTQEYQ	9	P22681:Q13191	
TM(+15.99)VGKTFTY	9	Q16850	Oxidation (M)
STYPRPHEY	9	Q9UBR2	
PTRAGRGH	8	Q92618	

APRQPGLM(+15.99)A	9	Q5T1J5:Q9Y6H1	Oxidation (M)
QLYDKGLVY	9	P41252	
KPAPAKPVAA	10	P12111	
RPWLEGRHTL	10	Q9UI10	
STVSRFFLY	9	Q969X6	
GYLHDFLKY	9	Q15014	
RPKETDHKSL	10	Q9BVI0	
NLLEKQLEY	9	Q86XR8	
HLYIQM(+15.99)EY	9	Q9P2K8	Oxidation (M)
SPAKNPSSL	9	Q9GZR7	
VLYDRVLY	9	Q9UHB9	
GNYEKVFEY	9	Q92624	
AEIVEGENHTY	11	P21333	
SPDIRNTTV	9	Q96PD2	
VFIVNSEDYM(+15.99)Y	11	O75182	Oxidation (M)
AEFKEAFQLF	10	P60660	
YFSDPFLKY	9	Q15165	
KLLPYIVGV	9	P12111	
FPNIPGKSL	9	O15121	
SPRLPVGGF	9	Q14669	
TPKM(+15.99)PGQSV	9	Q8WYP5	Oxidation (M)
KVLEILHRV	9	Q5GLZ8	
KLNPQQFEV	9	P46977	
YFIDRDGEIFRY	12	Q96S11	
FLASESLIKQI	11	P62906	
YPKRPLLGL	9	Q9GZR7	
KAYN(+.98)QTHLY	9	Q14563	Deamidation (NQ)
APRPDSRHSL	11	Q9BSV6	
LLLLLLRSAGCW	12	Q9UI38	
EEPTVIKKY	9	Q9Y5X3	
QM(+15.99)LEFAFRY	9	Q9HBM6:Q16594	Oxidation (M)
YVANEFLKY	9	Q14314	
ALLDRIVSV	9	Q92621	
YLRPPNTSL	9	O75494	
NPEKRYNVL	9	Q1KMD3	
EEHLESHKKY	10	Q9Y2D5	
SPNSKVNTL	9	P40939	
GVM(+15.99)DRGYSY	9	P28074	Oxidation (M)
EEAEIRKY	9	P09874	
ALSGHLETV	9	A6NMY6:P07355	

**Tableau 3** - Immunopeptidome des cellules MRC5 infectées (protéome viral)

Peptide	Lenght	Source	PTM
DTSGFAVYVKS KVG	14	Q01455	
HFSYVPTKY	9	P36334	
TSGFAVYVKS KVG	13	Q01455	
SGFAVYVKS KVG	12	Q01455	
SGFAVYVKS K	10	Q01455	
NINPHNYSY	9	P30215	
TSGFAVYVKS KVGNYR	16	Q01455	
FINSM(+15.99)QFNY	9	Q80872	Oxidation (M)
TADEAIKF	8	Q01455	