Université de Montréal

Développement de méthodes analytiques pour la protéomique et l'identification de peptides MHC I issus de cellules leucémiques

Par
Marie-Hélène Fortier

Département de Chimie
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Études Supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en chimie

Octobre 2008

Université de Montréal
Faculté des Études Supérieures

Cette thèse intitulée :

Développement de méthodes analytiques pour la protéomique et l'identification de peptides MHC I issus de cellules leucémiques

Présentée par:
Marie-Hélène Fortier

A été évaluée par un jury composé des personnes suivantes:

Karen Waldron ..... Président-rapporteur
Pierre Thibault ..... Directeur de recherche
Jean-François Masson ..... Membre du jury
Georg Holländer ..... Examinateur externe
Pascale Legault ..... Représentant du doyen

Sommaire

Les peptides présentés par les molécules MHC (Major Histocompatibility Complex) I jouent un rôle important au niveau de la régulation du développement des cellules T et de l'immunosurveillance. Certains peptides surexprimés à la surface des cellules leucémiques peuvent être utilisés comme cibles potentielles en immunothérapie. Le ciblage spécifique de ces peptides est présentement limité par nos connaissances fragmentaires concernant la diversité de la population allogénique présentée à la surface des cellules leucémiques.

Étant donné que les peptides MHC I sont présents en quantité infime à la surface des cellules, l'évaluation de différents systèmes chromatographiques fut effectuée pour tenter d'augmenter la sensibilité et la sélectivité de la séparation. Les avantages et les limitations d'un nouveau système microfluidique en terme de capacité, de reproductibilité et de sensibilité furent comparés aux performances analytiques des colonnes capillaires. Ces études ont permis d'établir que les deux systèmes étudiés sont comparables en terme de limite de détection ainsi qu'au niveau de la reproductibilité des intensités et des temps de rétention.

Une nouvelle approche analytique combinant différentes stratégies bioinformatiques aux techniques analytiques de séparation par 2D-nanoLC-MS fut proposée pour définir précisément le répertoire peptidique MHC I (MIP) de cellules normales et leucémiques. Cette étude a mené à l'identification de ~200 peptides présentés par des cellules normales et leucémiques. L'intégration des données peptidiques à celle du transcriptome du thymus a mené à la conclusion que le répertoire MIP de cellules normales thymiques contenait une signature spécifique de son tissus d'origine représentant ~17% des gènes du répertoire MIP. De plus, les résultats ont mis en évidence que le répertoire MIP des thymocytes était enrichi en peptides dérivant de transcrits très abondants et de protéines sources provenant de la famille des cyclines et des hélicases. La comparaison des répertoires MIP des cellules normales et leucémiques a permis de constater des variations significatives d'expression pour ~25% des peptides, dont plus de la moitié provenaient de protéines sources impliquées au niveau de la transformation néoplasique. Les essais d'immunisation effectués sur deux peptides surexprimés chez les cellules néoplasiques ont mené à la conclusion que des

changements d'expression de l'ordre de 10 à ≥ 85 fois pouvaient stimuler une réponse cytotoxique spécifique.

Cette méthodologie analytique fut aussi employée pour étudier la cinétique de présentation des peptides MHC I suite à l'inhibition de la voie oncogénique mTOR par la rapamycine, et a permis l'identification ~600 peptides présentés par des cellules leucémiques. La comparaison des profils cinétiques MHC I a révélé une augmentation progressive de l'expression de 70% du répertoire MIP après une exposition prolongée à la rapamycine, celle-ci corrélant avec une augmentation de l'expression de surface des molécules MHC I. De ce nombre, 17% des peptides sont significativement surexprimés et une grande proportion de ceux-ci sont associés à des protéines sources impliquées au niveau de la prolifération et de la progression du cycle cellulaire.

Ces deux études mettent donc en valeur l'importance de l'information unique générée par les analyses à haut débit MS, permettant une meilleure compréhension de la genèse du répertoire MIP.

Mots-clés: peptides MHC I, spectrométrie de masse, profil d'expression, microfluidique, chromatographie liquide capillaire, leucémie, transcriptome, immunothérapie, rapamycine, voie de signalisation mTOR.

Abstract

Peptides presented by MHC (Major Histocompatibility Complex) I molecules play vital roles in the regulation of T cell development and immunosurveillance. Peptides overexpressed at the cell surface of leukemic cells can be used as potential targets in cancer immunotherapy. However, the specific targeting of these peptides is actually hampered by our limited knowledge concerning the diversity of the allogenic population presented at the cell surface of leukemic cells.

Since MHC I peptides are present at low-level at the cell surface, different chromatographic systems were evaluated to enhance the sensitivity and the selectivity of the separation. The advantages and the limitations of a novel microfluidic system in terms of loading capacity, reproducibility and sensitivity were compared with the analytical performances of standard capillary columns. These studies showed that the limits of detection as well as the level of reproducibility of intensities and retention time are comparable using both systems.

A novel analytical approach combining different bioinformatics strategies with the analytical power of separation by 2D-nanoLC-MS was proposed to provide a molecular definition of the MHC class I peptide (MIP) repertoire of normal and leukemic cells. This study led to the identification of ~200 peptides presented by normal and leukemic cells. By integrating global profiling of the mouse protein-encoding transcriptome with the MIP repertoire of thymocytes, we found that the thymocytes' MIP repertoire is enriched in peptides derived from highly abundant transcripts and in peptide-source proteins derived from cyclins, cyclin-dependent kinases and helicases. Furthermore, our data suggest that the repertoire of MHC I-associated peptides conceals a tissue-specific signature that derives from ~17% of the genes represented in the MIP repertoire. The comparison of the MIP repertoire of normal and leukemic cells revealed that ~25% of peptides were differentially expressed. Interestingly, more than half of those peptides were derived from proteins implicated in neoplastic transformation. Immunization assays performed using two peptides shown to be overexpressed on neoplastic cells indicated that changes in expression of 10 to ≥ 85 fold can elicit specific cytotoxic T-cell responses.

This analytical approach was also used to monitor the interplay between the MIP repertoire and the mTOR oncogenic pathway following its inhibition with rapamycin, and enabled the identification ~600 peptides presented by leukemic cells. The kinetic profiles of MHC I peptides from control and rapamycin-treated EL4 cells indicated a progressive increase in abundance of 70% of the MIP repertoire, which also correlates with an increase level of cell surface expression of MHC I molecules. Approximately 17% of these peptides were differentially expressed following prolonged mTOR inhibition, and encoded a large proportion of source proteins associated with cell cycle progression and cellular proliferation.

Finally, these two studies highlight the importance of the unique information generated by high-throughput MS analyses, allowing a better understanding of the genesis of the MIP repertoire.

Keywords: MHC I peptides, mass spectrometry, expression profiling, microfluidic, capillary liquid chromatography, leukemia, transcriptome, immunotherapy, rapamycin, mTOR signaling pathway.

TABLE DES MATIÈRES

Liste des tableaux

Liste des figures

Liste des abréviations

2D-LC	Two-dimensional-liquid chromatography (Chromatographie liquide bidimensionnelle)
2D-nanoLC-MS	2D liquid chromatography nanoelectrospray mass spectrometry (2D chromatographie liquide couplée à la spectrométrie de masse par nanoélectronébulisation)
4E-BP1	eif4E-binding protein 1
ACN	Acetonitrile
AKT	Akt protein kinase
β_2m	β_2-microglobulin (β_2-microglobuline)
C_{18}	Octadecylsilane stationary phase (Phase stationnaire d'octadécylsilane)
Cdkn1b	Cyclin-dependent kinase inhibitor 1B
CE-MS	Capillary electrophoresis-mass spectrometry (Électrophorèse capillaire couplée à la spectrométrie de masse)
CFSE	Carboxyfluorescein succinimidyl ester
CID	Collision-induced dissociation (Dissociation induite par collision)
C-trap	C-shaped storage trap
Da	Dalton
DB	Database
DCs	Dendritic cells
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid (Acide désoxyribonucléique)
DRiPs	Defective ribosomal products
E/T	Effector-to-target ratio

EASE	Expression analysis systematic explorer software
ECD	Electron capture dissociation (Dissociation par capture d'électrons)
ER	Endoplasmic reticulum (Reticulun endoplasmique)
ERAP1	Endoplasmic reticulum aminopeptidase 1
ESI	Electrospray ionization (Ionisation par électronébulisation)
ETD	Electron transfert dissociation (Dissociation par transfert d'électrons)
FA	Formic acid (Acide formique)
FACS	Fluorescence-activated cell sorting
FBS	Fetal Bovine Serum
FE	Fold enrichment
FITC	Fluoresceine isothiocyanate
FT-ICR	Fourier Transform Ion Cyclotron Resonance (Résonance cyclotron des ions à transformée de Fourier)
GO	Gene ontology
GPI	Glycosylphosphatidylinositol
He	Helium
HLB	Hydrophilic lipophilic balanced
HPLC	High Performance Liquid Chromatography (Chromatographie liquide à haute performance)
HRP	Horseradish peroxidase
i.d.	Inner diameter
ID	Identification
IGF	Insulin growth factor

IgG	Immunoglobulin G
IL	Isotopic labeling (Marquage isotopique)
IP	Immunopurification
IPI	International Protein Index
IT	Ion trap (Trappe ionique)
KO	Knock out
LC3	Microtubule-associated protein 1 light chain 3
LC-MS	Liquid chromatography-mass spectrometry (Chromatographie liquide couplée à la spectrométrie de masse)
LIT	Linear ion trap (Trappe ionique linéaire)
LOD	Limit of detection
LTQ	Linear trap quadrupole (Trappe linéaire quadrupôle)
m/z	Mass-to-charge ratio (Ratio masse/charge)
MAE	Mild acid elution
MALDI	Matrix-assisted laser desorption ionization (Ionisation par désorption laser avec matrice)
MeOH	Methanol
MHC I	Class I Major histocompatibility complex (Complexe majeur d'histocompatibilité de classe I)
MHC II	Class II Major histocompatibility complex (Complexe majeur d'histocompatibilité de classe II)
MIP	MHC class I peptide (Peptide MHC de classe I)
MRM	Multiple reaction monitoring (Suivi de réactions multiples)

mRNA	Messenger ribonucleic acid (Acide ribonucléique messager)
MS	Mass spectrometry (Spectrométrie de masse)
MS/MS	Tandem mass spectrometry (Spectrométrie de masse en tandem)
mTOR	Mammalian target of rapamycin
mTORC1	mTOR complex 1
mTORC2	mTOR complex 2
Myb	Myeloblastosis oncogene
Na_2PO_4	Sodium phosphate dibasic
n_c	Peak capacity
NCBInr	National Center for Biotechnology Information nonredundant
NL	Neutral loss (Perte de fragments neutres)
o.d.	Outer diameter
P	P-value
p	Phosphorylation
p70S6K	Ribosomal protein S6 kinase
PCR	Polymerase chain reaction
PI3K	Phosphoinositide 3-kinase
PLC	Peptide-loading complex
ppm	Parts per million (parties par million)
Q-q-Q	Triple quadrupole hybrid (Hybride triple quadrupôle)
Q-q-TOF	Quadrupole time-of-flight hybrid (Hybride quadrupôle temps d'envol)

qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RDPs	Rapidly degraded proteins
Rfc2	Replication factor subunit 2
RICTOR	Rapamycin-insensitive companion of mTOR
RP	Reversed phase (Phase inverse)
RSD	Relative standard deviation
Rt	Retention time (Temps de retention)
S/N	Signal-to-noise ratio (Ratio signal/bruit)
SCX	Strong cation exchange (Échange par cations forts)
SD	Standard deviation
Sgk	Serum/glucocorticoid regulated kinase
sMHC	Soluble MHC molecules (Molécules MHC solubles)
smm	Stabilized matrix method
$t_{1/2}$	Half-life
TAAs	Tumor associated antigens (Antigènes associés aux tumeurs)
TAP	Transporter associated with antigen processing
TBS	Tris-buffered saline
Tnfsf10	Tumor necrosis factor superfamily member 10
TOF	Time-of-flight (Temps d'envol)
TRAIL	TNF-related apoptosis inducing ligand
UV	Ultraviolet

WT	Wild type
Xlr3	X-linked lymphocyte-regulated 3
z	Charge

À mon cher Hugo

À ma petite fille Leyna

À ma famille

Remerciements

Je voudrais remercier mon directeur de recherche Dr. Pierre Thibault pour m'avoir accueillie dans son laboratoire et de m'avoir donnée l'opportunité de travailler sur un projet de thèse multidisciplinaire aussi stimulant. Merci pour ses encouragements et ses nombreux conseils ainsi que pour sa patience et sa compréhension lors des moments plus difficiles.

Mes remerciements vont ensuite à mes collaborateurs de l'IRIC, Etienne Caron et Dr. Claude Perreault, pour leur contribution respective à l'élaboration des concepts et des expériences biologiques. Merci pour leur enthousiasme, leur positivisme ainsi que pour leurs encouragements tout au long de mon doctorat. Je voudrais remercier aussi mes autres collaborateurs Dr. Sébastien Lemieux, Grégory Voisin et Marie-Pierre Hardy ainsi que les employés des différentes plateformes de l'IRIC qui ont contribué au succès de ce projet.

Je voudrais remercier le Dr. Éric Bonneil pour son aide et ses nombreux conseils par rapport à l'instrumentation LC-MS. Son support technique fut grandement apprécié tout au long de mon doctorat. Je tiens aussi à remercier Gagandeep Jaitly, Kevin Eng ainsi que Mathieu Courcelles pour leur assistance technique dans l'analyse des données à l'aide des outils bioinformatiques. Merci aussi à toute l'équipe de recherche du Dr. Pierre Thibault ainsi qu'à celle du Dr. Claude Perreault pour leurs conseils et leur support moral pendant toute la durée de mon doctorat.

J'aimerais remercier l'équipe d'Agilent Technologies et plus particulièrement Dr. Paul Goodley, Dr. Michnowicz ainsi que Georges Gauthier pour leur assistance dans le projet chipLC. Je remercie aussi Linda Côté et Sébastien Marchand pour leur support technique.

Je remercie le Conseil de recherche en sciences naturelles et en génie du Canada (CRSNG), le Fonds de la recherche en santé du Québec (FRSQ) ainsi que la Faculté des études supérieures (FES) pour leur support financier.

Finalement, merci à mon conjoint Hugo qui m'a donnée la force de continuer même dans les moments difficiles. Sa patience, ses encouragements ainsi que son support moral furent

énormément appréciés tout au long de mes études. Merci à ma famille ainsi qu'à la gang du dojo de Montréal Karaté Kenpo Jean-Guy Angell pour m'avoir toujours encouragée dans la poursuite de mes objectifs.

1. Introduction générale

1.1 Le Cancer

Le cancer est une maladie génétique caractérisée par une prolifération cellulaire anormale et anomique au sein d'un tissu ou d'une lignée cellulaire normale de l'organisme [1]. Cette maladie est la première cause de mortalité au Québec depuis 2005 et deviendra d'ici peu la principale cause au Canada. On estime à 166 400 le nombre de nouveaux cas de cancer et à 73 800 le nombre de décès par cette maladie qui surviendront au Canada en 2008 [2]. D'après les taux d'incidence actuels, plus d'un Canadien sur trois développera un cancer au cours de sa vie et un Canadien sur quatre mourra de cette maladie. On estime que d'ici quelques années, ce sera une personne sur deux qui souffrira d'au moins un cancer au cours de sa vie [2].

1.1.1 Effets biologiques sur la cellule

Le cancer résulte de mutations des cellules somatiques. La plupart des cancers dérivent d'une seule cellule affectée par une ou des mutations lui conférant un avantage sélectif. Ces nouvelles caractéristiques biologiques lui permettent de se diviser plus vigoureusement que ses voisines pour fonder un clone de mutants [1]. La genèse d'un cancer nécessite typiquement que plusieurs accidents indépendants et rares se produisent dans la lignée de la cellule. Les facteurs de risques sont soit internes (e.g. génome, mutation) ou peuvent être induits par des sources externes comme des agents chimiques, des virus et diverses formes de radiations [1].

Trois grandes catégories de gènes sont associées aux pathologies cancéreuses : les oncogènes, les gènes suppresseurs de tumeurs et les gènes impliqués dans la réparation de l'ADN [3]. Une augmentation de la prolifération cellulaire est observée dans les cellules atteintes en fonction de l'hyperactivation (oncogènes) ou de l'inactivation (suppresseurs) de certains gènes cibles. Ces évènements génétiques mènent à une cascade d'évènements biologiques affectant entre autre la synthèse des protéines. Les altérations génétiques peuvent aussi être associées à des anomalies plus importantes comme dans le cas des translocations chromosomiques [4]. Ces réarrangements chromosomiques sont à la source

de 20% des cancers et sont la principale cause du développement des leucémies et des lymphomes [5].

Une cellule doit donc acquérir une gamme complète de propriétés anormales pour engendrer le cancer. On retrouve sept caractéristiques capitales qui permettent aux cellules une croissance cancéreuse : i) elles fournissent leurs propres signaux de croissance, ii) elles ignorent les signaux inhibiteurs de la croissance cellulaire, iii) elles se divisent sans limite, iv) elles contrôlent leur angiogenèse, v) elles échappent aux mécanismes d'apoptose, vi) elles sont envahissantes et prolifèrent dans des sites étrangers et vii) elles échappent à l'immunosurveillance [6, 7].

1.1.2 Implication de la voie de signalisation PI3K-AKT-mTOR

Plusieurs voies de signalisation cellulaire furent identifiées comme jouant un rôle crucial au niveau de la progression du cancer. Parmi celles-ci, on compte la voie de signalisation PI3K-AKT-mTOR (mTOR, mammalian Target Of Rapamycin) qui est présentement une des plus étudiées [8-11]. Cette voie de signalisation joue un rôle essentiel dans la régulation de processus tels la synthèse des protéines, la biogenèse des ribosomes, l'autophagie ainsi que dans la progression du cycle cellulaire [9, 11-13]. Ces processus biologiques sont critiques au niveau de la prolifération, de la croissance et de la différentiation cellulaire ce qui confère un avantage significatif à la progression tumorale. Il n'est donc pas surprenant que la voie de signalisation PI3K-AKT-mTOR soit la plus fréquemment amplifiée dans les cellules cancéreuses [8, 14].

La figure 1.1 montre les principales composantes de la voie de signalisation PI3K-AKT-mTOR. En réponse aux facteurs de croissance, la kinase PI3K est activée et il s'en suit une cascade d'évènements menant à l'activation des kinases AKT et mTOR (Figure 1.1). La kinase mTOR est la composante centrale de cette voie car elle régule la croissance cellulaire et la synthèse protéique en contrôlant l'activité des kinases S6K et 4E-BP1. mTOR est donc une cible potentielle dans le traitement de différentes pathologies comme le cancer, le diabète, l'obésité, les désordres neurologiques et les maladies cardiovasculaires [15]. Il est d'ailleurs possible d'inhiber spécifiquement mTOR à l'aide de la drogue rapamycine et ses dérivés [15, 16]. Des essais cliniques sont d'ailleurs présentement en

cours sur différents types de cancers et montrent des résultats prometteurs pour le traitement de carcinomes rénaux ainsi que pour certaines leucémies et lymphomes [16, 17].

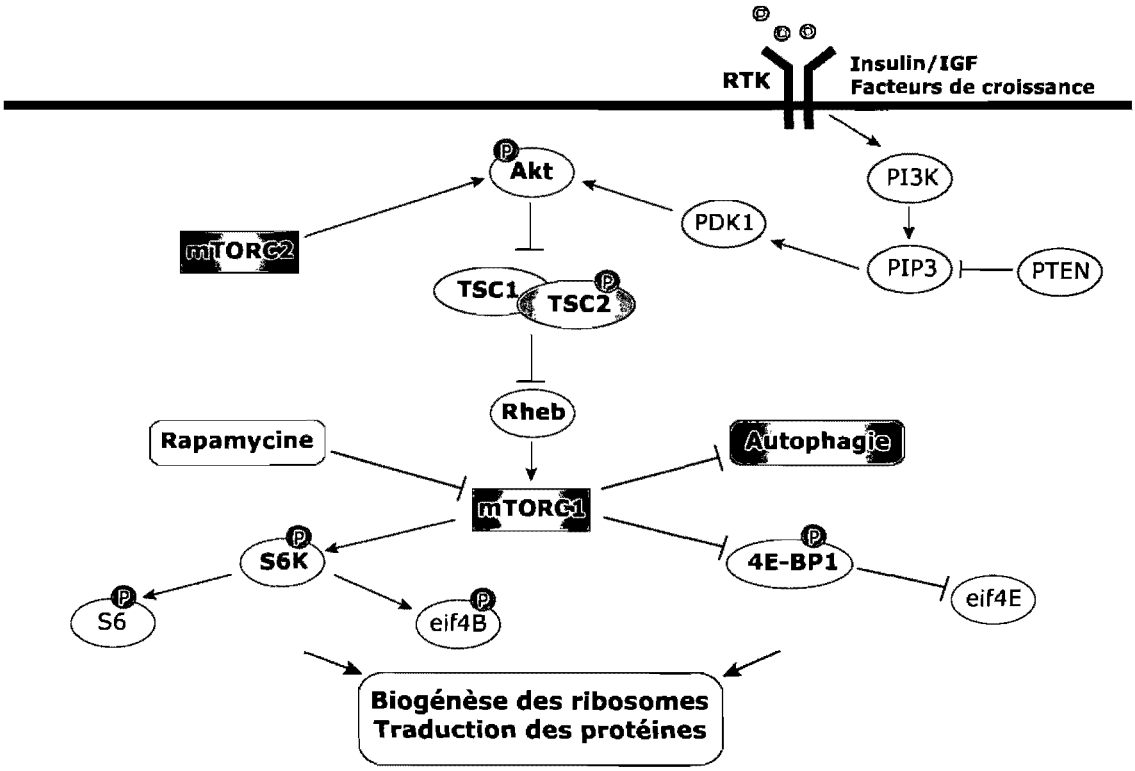

Figure 1.1: La voie de signalisation PI3K-AKT-mTOR. Flèche : activation; Barre : Inhibition; P : phosphorylation.

1.1.3 Traitement du cancer

Les cancers sont présentement traités par la combinaison de différentes approches comme la chirurgie, la chimiothérapie et la radiothérapie locale (e.g. irradiation de la tumeur par rayonnement X) [3]. Cependant, même quand le cancer semble apparemment vaincu, certaines tumeurs dormantes mènent fréquemment à des rechutes et à un échec thérapeutique. Pour contrer le cancer, les chercheurs tentent de développer des stratégies permettant de détruire efficacement toutes les cellules cancéreuses dans l'organisme. Il est de plus en plus évident que le système immunitaire a un rôle crucial à jouer dans le développement et l'éradication du cancer [7]. Des nouvelles approches thérapeutiques basées sur l'immunothérapie furent donc proposées pour détruire sélectivement les cellules cancéreuses par l'intermédiaire du système immunitaire [18]. Le développement d'applications cliniques combinant le pouvoir de l'immunothérapie à celui des techniques

actuelles est une approche des plus prometteuses pouvant permettre d'améliorer le taux de succès face au traitement du cancer [3, 19]. Une autre option consiste à utiliser le système immunitaire pour développer des vaccins contre le cancer [20].

1.2 Immunité adaptative

L'immunité adaptative consiste en un mécanisme de défense sophistiqué pour cibler ce qui est perçu comme une menace pour l'organisme. Cette réponse adaptative est hautement spécifique et possède une mémoire immunitaire [21]. Les réponses immunitaires adaptatives sont assumées par les lymphocytes qui reconnaissent les micro-organismes de façon spécifique, que ceux-ci soient situés à l'intérieur des cellules de l'hôte ou à l'extérieur dans les liquides biologiques [1, 21]. Il existe trois grandes classes de cellules impliquées dans la réponse immunitaire adaptative : les lymphocytes B qui marquent les agents étrangers par la production d'anticorps, les lymphocytes T auxiliaires (CD4) qui facilitent l'activation des lymphocytes B et les lymphocytes T cytotoxiques (CD8) qui tuent directement les cellules infectées en induisant chez elles une apoptose [1, 21]. La reconnaissance des antigènes par les lymphocytes T se fait seulement quand ceux-ci sont présentés à la surface d'autres cellules par les molécules du complexe majeur d'histocompatibilité (MHC). Ces antigènes sont des fragments peptidiques liés de façon non-covalente aux molécules du MHC. Les lymphocytes T CD4 reconnaissent des peptides étrangers liés aux molécules MHC de classe II (MHC II) provenant de la voie de présentation endosomale. En comparaison, les lymphocytes T CD8 réagissent face à des peptides étrangers associés aux molécules MHC de classe I (MHC I) dérivant de la voie de présentation cytosolique [22, 23]. Les molécules du MHC I sont exprimées à la surface de toutes les cellules nucléées de l'organisme tandis que celles du MHC II sont retrouvées majoritairement sur les cellules présentatrices d'antigènes comme les macrophages, les lymphocytes B et les cellules dendritiques [1].

1.2.1 Les peptides du MHC I

Le complexe peptide-MHC I est composé d'une chaîne lourde glycosylée étant associée de manière non-covalente à la protéine β_2-microglobuline (β_2m) et au peptide

antigénique [21]. Les peptides présentés par les molécules MHC I sont de courtes chaînes peptidiques de 8 à 11 acides aminés qui dérivent de la dégradation des protéines intracellulaires du soi ou du non-soi (protéines étrangères). On retrouve donc des milliers de peptides MHC I à la surface d'une même cellule dont l'abondance varie de l'ordre du 1 à 10 000 molécules/cellule [24, 25]. Les peptides sont ancrés dans la cavité des molécules MHC I par l'intermédiaire de résidus spécifiques (motif peptidique). Les protéines membranaires MHC I sont connues pour être hautement polymorphiques, chacune des classes alléliques présentant son motif d'ancrage spécifique [22, 24, 26]. On retrouve donc un niveau d'homologie structurale très élevé au sein d'une même population allélique. Cette caractéristique fut d'ailleurs exploitée dans le cadre du développement d'algorithmes de prédiction, permettant entre autre l'évaluation de l'affinité des peptides pour les différentes classes de molécules MHC I [27-29]. Dans les études présentées dans cette thèse, des cellules provenant de souris C57BL/6 furent utilisées comme modèle cellulaire. Ces souris C57BL/6 présentent des molécules MHC I classiques de type $H2D^b$ et $H2K^b$ ainsi que des molécules MHC I non classiques de type Qa1 et Qa2. Les motifs peptidiques des principaux allèles présentés dans cette thèse sont illustrés dans la figure 1.2.

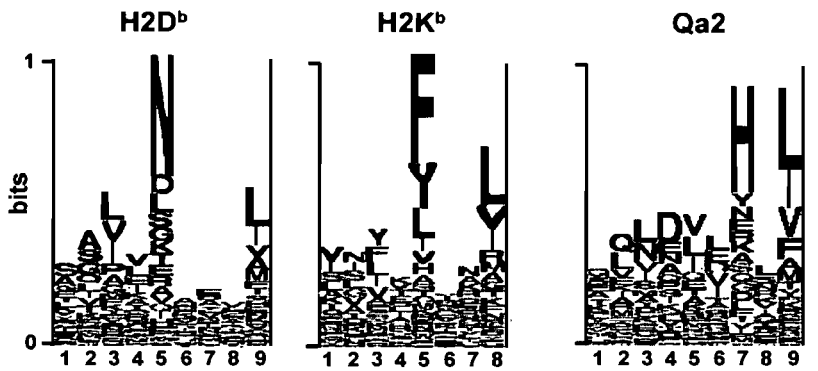

Figure 1.2: Motifs peptidiques MHC I présentés chez des souris de type C57BL/6. Les logos illustrent le profil des motifs peptidiques pour les molécules de type $H2D^b$, $H2K^b$ et Qa2. Les acides aminés acides (rouge), basiques (bleu) et hydrophobiques (noir) sont représentés. Modifiée de [30].

1.2.1.1 La présentation des peptides par les molécules du MHC I

Le protéasome joue un rôle crucial au sein des mécanismes d'immunosurveillance [31]. Lorsqu'une protéine étrangère est présentée dans l'environnement intracellulaire ou

qu'une erreur survient dans la traduction d'une protéine (Figure 1.3, point 1), celle-ci est marquée par une chaîne polyubiquitine (Figure 1.3, point 2) étant reconnue par le protéasome. Les protéines ubiquitinées sont ensuite dégradées par le protéasome en chaînes peptidiques (Figure 1.3, point 3) [31]. Une certaine proportion de ces peptides est ensuite exportée dans le réticulum endoplasmique (ER) par l'intermédiaire du canal TAP (Transporter associated with Antigen Processing) (Figure 1.3, point 4). Au sein du ER, ces peptides sont remodelés par l'aminopeptidase ERAP1 [32] et les peptides montrant une affinité spécifique sont incorporés dans les molécules du MHC I via le complexe de chargement des peptides (Figure 1.3, point 5) [33, 34] permettant ainsi de former des complexes peptide-MHC I stables pouvant être exportés à la surface de la cellule (Figure 1.3, point 6) [35]. Ce mécanisme de présentation est un cycle continuel permettant de renouveler le répertoire peptidique présenté à la surface des cellules. Toute modification dans l'environnement intracellulaire, causée par une instabilité génomique ou par l'intrusion d'un agent pathogène, provoquera donc des changements au sein de la population peptidique MHC I. Ces modifications pourront ensuite être perçues par le système immunitaire. Malgré que certaines exceptions subsistent toujours, il est maintenant largement accepté que la majorité des peptides sont générés et présentés par cette voie de présentation [36].

1.2.1.2 La source de ces peptides antigéniques

L'ensemble du protéome peut contribuer à la génération de peptides antigéniques. Cependant, étant donné qu'on retrouve seulement environ 100 000 molécules MHC I à la surface d'une même cellule et que le protéome est constitué d'environ 30 000 protéines différentes, plusieurs d'entre elles ne pourront être détectées par le système immunitaire [37]. Les protéines peuvent être dégradées à différentes vitesses suite à leur synthèse : certaines présentent des temps rapides de dégradation de l'ordre des minutes tandis que d'autres se dégradent très lentement sur plusieurs jours [38]. De plus, on estime que seulement 0,1% des peptides spécifiques survivent à la dégradation intracellulaire et sont présentés à la surface des cellules. La question suivante peut donc être posée : « En tenant compte de ces différents aspects, comment peut-on expliquer que des peptides antigéniques dérivant de certaines protéines virales connues pour se dégrader lentement peuvent être présentés au système immunitaire moins d'une heure suivant l'infection? » L'hypothèse la

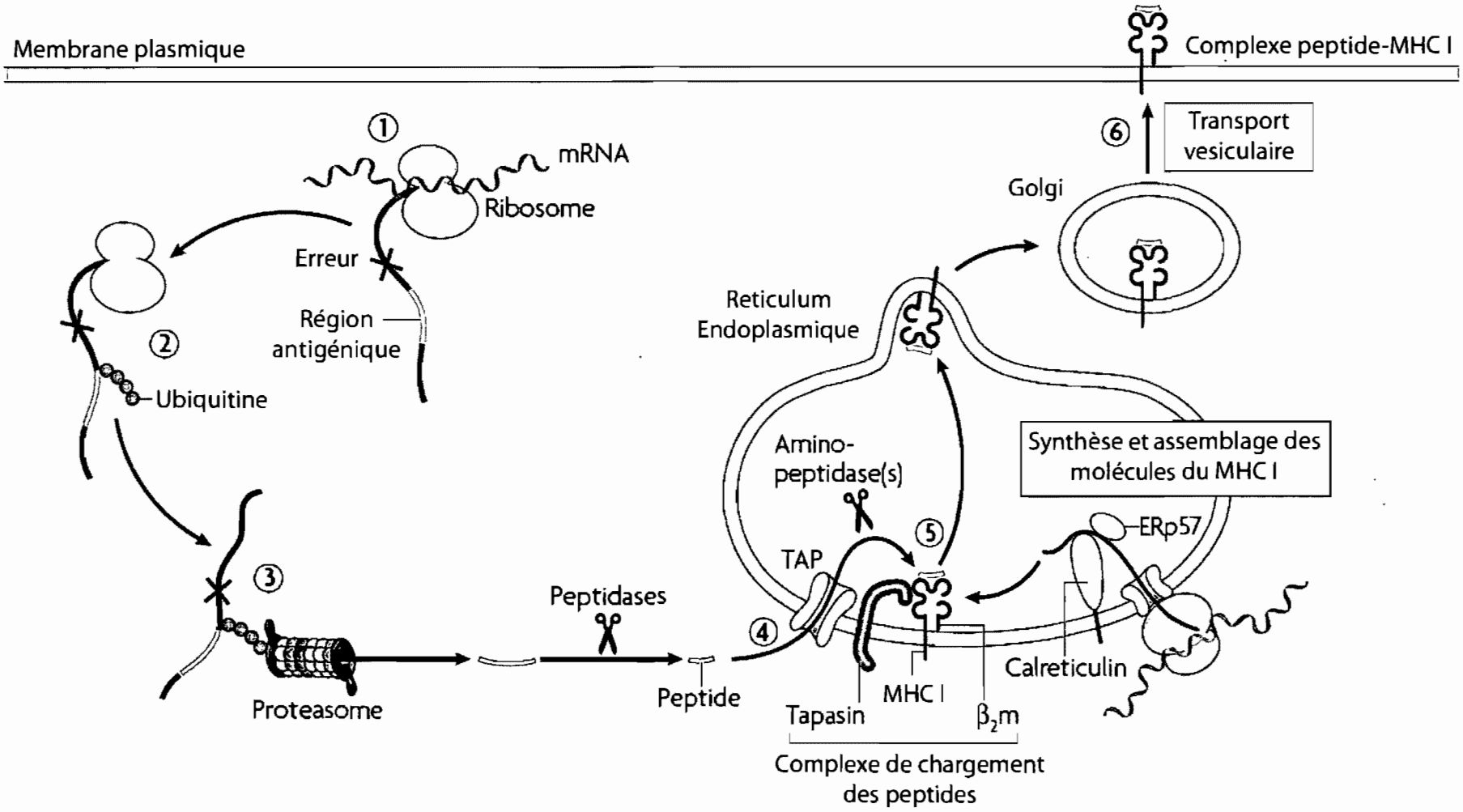

Figure 1.3: Mécanisme de la présentation antigénique par les molécules du MHC I. Modifiée de [36].

plus acceptée jusqu'à présent est celle des DRiPs (Defective Ribosomal Products) proposant qu'une grande proportion des peptides présentés par les molécules du MHC I proviendrait de protéines défectueuses nouvellement synthétisées n'ayant jamais atteint leur conformation native [37-40]. Des études récentes ont d'ailleurs mis en évidence l'existence réelle des DRiPs ainsi que leur contribution majeure à la génération des peptides antigéniques [39, 41]. Ces études ont démontré qu'une proportion importante de peptides antigéniques dérivaient de protéines nouvellement synthétisées [41] et qu'environ 30% d'entre elles dérivaient des DRiPs, lesquelles sont rapidement dégradées par le protéasome (Figure 1.4) [39]. Bien que les résultats répertoriés jusqu'à présent semblent indiquer que l'hypothèse des DRiPs reste la plus probable, celle-ci suscite encore des discussions face à l'exactitude du modèle [42].

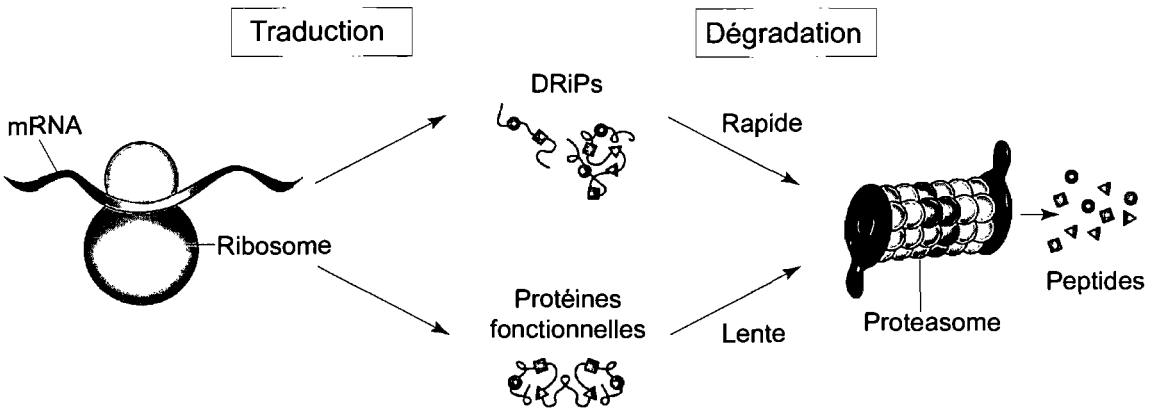

Figure 1.4: La source des peptides antigéniques. Modifiée de [37].

1.2.1.3 Implications de la présentation antigénique au niveau du cancer

Les mutations génétiques prenant place au sein des cellules suite à la transformation néoplasique peuvent changer la composition du répertoire peptidique MHC I présenté au système immunitaire. Certains antigènes associés aux tumeurs (TAAs) ont été identifiés sur différents types de cellules cancéreuses [43]. Le ciblage des potentiels TAAs est par contre difficile, car les cellules néoplasiques ont développé différents mécanismes biologiques pour échapper à l'immunosurveillance [7]. Présentement, les chercheurs cherchent toujours le moyen d'exploiter le pouvoir du système immunitaire pour détruire sélectivement les cellules cancéreuses dans l'organisme. Des approches immunothérapeutiques furent

développées pour aider le système immunitaire à cibler les cellules néoplasiques [23]. Ces approches furent proposées comme traitement et pour l'immunisation contre le cancer [20]. Cependant, pour développer ces techniques, les immunologistes ont besoin de connaître les structures moléculaires des TAAs et utiliser une approche efficace pour déterminer quels candidats permettraient une réponse des plus sélectives. Le polymorphisme des molécules du MHC I ainsi que le manque de connaissances face à la composition des peptides antigéniques présentés par les cellules cancéreuses sont les facteurs importants limitant actuellement le développement de ces thérapies. La protéomique se présente donc comme un outil indispensable pour résoudre la problématique associée à l'étude du répertoire peptidique présenté par les cellules cancéreuses.

1.3 La spectrométrie de masse en protéomique

La protéomique est une approche expérimentale permettant l'étude des processus biologiques cellulaires par l'analyse systématique des protéines exprimées dans des cellules ou des tissus [44]. Cette science est complémentaire à la génomique (étude du génome) et est de plus en plus utilisée dans tous les domaines de la biologie (e.g. biomédical) que ce soit pour l'étude des profils protéiques globaux ou pour la quantification des changements à travers différents protéomes. La protéomique est devenue essentielle à la compréhension des mécanismes cellulaires face à la grande diversité et la complexité des protéomes (e.g. polymorphisme séquentielle, modifications post-traductionnelles) [45]. Le développement de techniques efficaces pour l'ionisation douce des biomolécules par spectrométrie de masse (MS) comme l'électronébulisation (ESI) [46] et l'ionisation par désorption laser avec matrice (MALDI) [47] ont propulsé de manière significative l'étude à grande échelle du protéome depuis le début des années 1990. La spectrométrie de masse est devenue une technique de choix pour l'analyse des mélanges complexes protéiques fournissant des réponses uniques à la compréhension de la biologie des systèmes et aux mécanismes moléculaires régulant le fonctionnement de la cellule. L'utilisation de la spectrométrie de masse en protéomique offre une vaste gamme d'applications incluant le catalogage du protéome d'organismes et d'organelles, l'analyse différentielle des protéines exprimées dans différentes populations cellulaires ou conditions expérimentales, l'étude des réseaux d'interaction protéiques ainsi que la caractérisation des modifications post-traductionnelles [48].

1.3.1 Les méthodes de séparation

Les techniques employant la chromatographie liquide couplée à la spectrométrie de masse (LC-MS) sont présentement les plus utilisées pour l'étude des mélanges complexes protéiques. La chromatographie liquide à haute performance (HPLC) offre un pouvoir de résolution considérable pour la séparation des peptides par la diversité des modes de séparation possibles ainsi que pour la vaste gamme de phases stationnaires disponibles [49]. Le couplage LC-MS permet donc d'avoir une technique de séparation versatile ainsi qu'une instrumentation des plus puissantes pour la détection et l'identification des biomolécules. La complémentarité au niveau de la séparation chromatographique, les capacités de chargement supérieures, ainsi que la facilité du couplage LC-MS en mode ESI ont propulsé cette approche analytique à l'avant-plan par rapport à l'électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS) qui est une technique beaucoup moins reproductible et robuste que la LC-MS [50].

Un des défis de la protéomique est d'être capable de détecter des protéines présentes en faible abondance dans des matrices biologiques extrêmement complexes. Pour répondre à ce besoin, des techniques telles que la nanoLC-MS furent développées permettant ainsi d'augmenter le pouvoir de résolution et la sensibilité (S/N, signal/bruit)[a] dans l'analyse d'extraits protéiques. Les approches analytiques par nanoLC-MS utilisent typiquement des colonnes capillaires en mode phase inversée (RP) de l'ordre du 50-150 µm en diamètre interne qui sont couplées en ligne au spectromètre de masse à l'aide d'interfaces nanoESI opérant à des débits très faibles de l'ordre du 100-600 nL/min permettant ainsi d'augmenter de façon significative la sensibilité analytique [51, 52]. La recherche constante d'améliorer les performances chromatographiques en nanoLC-MS a aussi mené au développement de systèmes microfluidiques compacts [53]. La configuration de ces puces nanoLC permet d'avoir une réduction significative des volumes morts par la diminution des lignes de transfert et des connexions permettant ainsi d'augmenter la sensibilité des analyses en protéomique [54]. Étant donné que les extraits biologiques sont rares et doivent être utilisés pour plusieurs analyses, la miniaturisation des systèmes est un grand défi en protéomique pour limiter la quantité de matériel nécessaire.

Pour permettre de résoudre la complexité associée à des extraits biologiques réels (e.g. lysats cellulaires) qui comportent des milliers de protéines différentes, des techniques

[a] Définition couramment utilisée en protéomique.

de séparation complémentaires ont dû être développées afin d'augmenter la capacité chromatographique en LC-MS. Une des méthodologies populaires consiste à fractionner les protéines sur des gels d'électrophorèse 1D ou 2D préalablement à la digestion trypsique et à l'analyse en LC-MS [55]. Une approche alternative au fractionnement des protéines est la séparation bidimensionnelle des peptides trypsiques combinant deux modes de séparation complémentaires comme la chromatographie d'échange d'ions (e.g. SCX, Strong Cation eXchange) et la chromatographie en phase inversée (RP) [56, 57]. La seconde dimension peut être faite par fractionnement (collection de fractions) ou bien être couplée en ligne avec le système LC-MS [58, 59]. Cette séparation bidimensionnelle permet d'augmenter le pouvoir de résolution grâce à la complémentarité des phases facilitant ainsi la dispersion de la population peptidique. Ces deux approches permettent donc de diminuer la complexité des extraits protéiques et ainsi augmenter la capacité de détecter des protéines présentées en faible abondance dans des matrices biologiques complexes.

1.3.2 La spectrométrie de masse

La spectrométrie de masse (MS) est devenue une technique analytique avantageuse pour l'analyse d'échantillons chimiques ou biochimiques car elle offre des informations importantes sur la structure primaire, la composition et la pureté des analytes. Le spectromètre de masse est un instrument séparant des analytes chargés en phase gazeuse selon leur ratio masse/charge (m/z) et mesurant l'abondance relative de chacun des ions présents dans l'échantillon. Ces informations sont ensuite collectées sous forme d'un spectre de masse [60]. La MS est présentement considérée comme un outil indispensable en protéomique et dans une multitude de disciplines. Parmi les avantages majeurs de l'utilisation de la MS en protéomique, on compte la haute sensibilité, l'exactitude de masse, l'efficacité de séparation ainsi que la capacité de générer de l'information structurale riche permettant l'identification efficace de peptides et de protéines.

1.3.2.1 L'ionisation par électronébulisation (ESI)

Pour une longue période de temps, la caractérisation par MS a été restreinte à l'analyse de petites molécules thermiquement stables étant donné le manque de techniques efficaces pour l'ionisation douce des composés de la phase liquide à gazeuse sans fragmentation excessive. Le développement des techniques d'ionisation douce ESI et

MALDI a rendu accessible l'analyse des biomolécules par spectrométrie de masse favorisant ainsi l'expansion exponentielle des approches protéomiques [46, 47]. L'intérêt général fut davantage marqué pour l'ionisation ESI, une technique permettant le couplage en ligne LC-MS et CE-MS [56].

En ESI, une solution d'analytes est introduite à faible débit à travers un capillaire qui est soumis à un potentiel élevé (2-5 kV). Le champ électrique charge le solvant contenant l'analyte à l'extrémité du capillaire produisant un fin pinceau de gouttelettes chargées (cône de Taylor, Figure 1.5). Les gouttelettes peuvent être chargées positivement ou négativement dépendant de la polarité du voltage. Par la suite, les gouttelettes diminuent en taille par un processus d'évaporation jusqu'au point où elles atteignent la limite de Rayleigh. À ce point, la densité de charge à la surface des gouttelettes est égale à la tension de surface du liquide et les forces électrostatiques élevées mènent à une explosion coulombique. Ce processus est répété jusqu'à ce que la diminution en taille des gouttelettes résulte en la formation d'espèces moléculaires multiplement chargées [61-63].

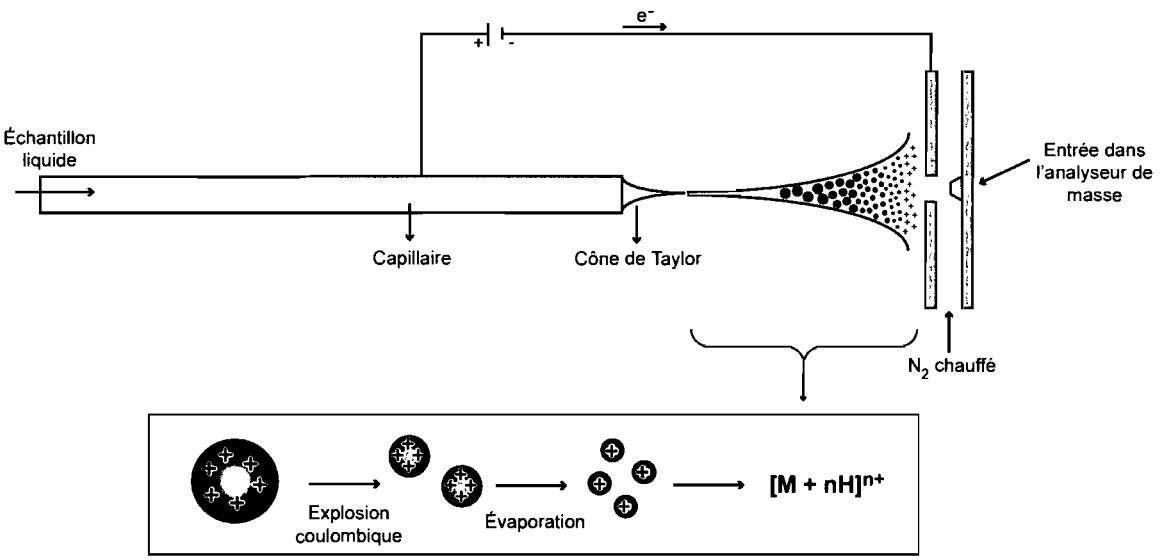

Figure 1.5: Mécanisme d'ionisation par électronébulisation (ESI, mode positif).

Pour augmenter la sensibilité (S/N) de cette technique d'ionisation, des interfaces nanoESI furent développées pour opérer à des débits dans l'ordre du nL/min. La diminution du débit et la configuration des sources nanoESI par rapport aux techniques ESI classiques permettent d'améliorer l'efficacité de l'ionisation, du processus d'évaporation ainsi que le transfert des ions dans le système sous vide de deux ordres de grandeur (~100X) [64]. Cette

technique d'ionisation est maintenant largement répandue dans l'analyse des biomolécules ainsi que pour la détection de petites molécules présentées à l'état de trace.

1.3.2.2 Les analyseurs de masse

Les analyseurs de masse sont à la base de la séparation par MS. Différents types d'analyseurs de masse existent présentement sur le marché chacun possédant un mécanisme qui leur est propre pour la séparation des ions selon leur ratio masse/charge (m/z). En protéomique, les analyseurs recherchés sont ceux qui permettent à la fois de faire de la détection MS ainsi que le séquençage peptidique par spectrométrie de masse en tandem (MS/MS). Différents instruments hybrides sont maintenant disponibles facilitant ainsi les applications en protéomique. Parmi les instruments les plus couramment utilisés en protéomique, on retrouve les trappes ioniques (IT) et les trappes ioniques linéaires (LIT), les triple quadrupôles (Q-q-Q), les hybrides quadrupôle temps d'envol (Q-q-TOF) ainsi que les instruments à haute résolution comme le spectromètre de masse à résonance cyclotron des ions à transformée de Fourier (FT-ICR) ou bien plus récemment l'hybride LTQ-Orbitrap [45, 56, 65]. Chacun des instruments possédant ses forces et ses faiblesses, le choix de l'appareil doit être fait de façon judicieuse selon le type d'analyse à effectuer. Le tableau 1.1 fait un résumé des caractéristiques et performances de chacun des instruments en fonction des différents paramètres analytiques à considérer [45].

Tableau 1.1: Caractéristiques et performances des spectromètres de masse les plus couramment utilisés en protéomique (en mode ESI-MS).

	IT/ LIT	Q-q-TOF	FT-ICR/ LTQ-Orbitrap	Q-q-Q/ Q-q-LIT
Exactitude de masse	+	+++	++++	++
Résolution	+	+++	++++	+
Sensibilité (S/N)	+++	++	++ à +++	+++
Gamme dynamique	+	++	++	+++
MS/MS	oui	oui	oui	oui
Options supplémentaires	MS^n, ETD		ECD/ MS^n, ETD	P, NL, MRM
Identification	+++	+++	++++	++
Quantification	++	++++	+++	++++
Débit (Throughput)	++++	+++	+++	+++

Les symboles +, ++, +++ et ++++ indiquent possible ou faible, moyen, bon ou élevé, et excellent ou très élevé, respectivement. ETD, Electron Transfer Dissociation; ECD, Electron Capture Dissociation; P, Precursor scan; NL, Neutral Loss scan; MRM, Multiple Reaction Monitoring

1.3.2.3 La spectrométrie de masse en tandem (MS/MS)

La spectrométrie de masse en tandem (MS/MS) est une technique utilisée pour obtenir de l'information sur la structure primaire des peptides présents dans un échantillon. En MS/MS, on isole un ion peptidique particulier dont le gain d'énergie interne suite à l'activation par collision entraîne la fragmentation de l'ion le long de sa structure peptidique. En mode de dissociation induite par collision (CID), le gain d'énergie interne survient suite aux collisions de l'ion avec des atomes de gaz inerte (azote, argon ou hélium) [57]. Ce mode de fragmentation est celui qui est le plus couramment utilisée dans les instruments de type Q-q-TOF, Q-q-Q et IT. En mode CID à faible énergie, la fragmentation est majoritairement observée aux liens amides des acides aminés étant donné que la rupture de ces liaisons covalentes requiert moins d'énergie. Le patron de fragmentation séquentielle le long de la chaîne peptidique dépendra entre autres de la composition en acides aminés basiques ainsi que de la mobilité du proton à l'extrémité N-terminale [66]. Selon la nomenclature établie, la rupture des liens amides permet de former des ions de type « b » et « y » (Figure 1.6) [66, 67]. Les spectres CID contiennent donc majoritairement des ions de la série « y » présentant la charge positive sur le résidu en C-terminal, des ions de la série « b » présentant la charge en N-terminal, ainsi que des ions « y » et « b » avec pertes de molécules d'eau ou d'ammoniaque. Il est aussi possible d'observer des fragments de type « a » dérivant de la perte de carbonyle (C = O) d'ions « b ». L'analyse du spectre MS/MS permet de déduire la séquence peptidique à partir des différences de masse observées entre les fragments consécutifs d'une même série.

L'analyse des spectres MS/MS peut être faite par séquençage *de novo* ou bien par la recherche dans les banques de données. Pour l'analyse par séquençage *de novo*, la structure de la protéine est inconnue ce qui requiert une interprétation manuelle des spectres avec ou sans l'aide de logiciels spécialisés (e.g PEAKS) [68]. L'approche *de novo* est donc très dépendante de la qualité des spectres en terme d'exactitude de masse sur les ions fragments, de la résolution de l'instrument ainsi que par rapport à la qualité du patron de fragmentation [57]. En comparaison, l'analyse des spectres MS/MS par la recherche dans les banques de données est basée sur le principe que les protéines sont connues et répertoriées. Cette méthode est davantage utilisée pour les analyses à haut débit étant donné ces avantages au niveau de la sensibilité et de la rapidité dans l'identification des protéines.

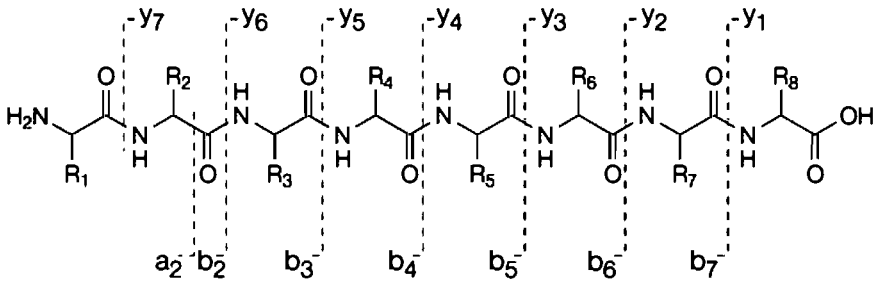

Figure 1.6: Fragmentation des peptides en mode CID. Reproduite de [57].

1.3.2.4 La recherche dans les banques de données

L'implantation des approches utilisant la recherche dans les banques de données fut un développement important pour maximiser l'identification à haut débit des protéines présentes dans des mélanges complexes. Le principe de ces approches consiste à établir une corrélation entre les spectres peptidiques expérimentaux MS/MS et les spectres théoriques présentés dans les banques de données (Figure 1.7) [69]. Différents algorithmes mathématiques comme celui de Mascot et Sequest, par exemple, furent développés pour calculer des pointages de corrélation entre les spectres expérimentaux et théoriques [57]. Ces pointages de corrélation sont évalués pour chaque spectre expérimental, et le pointage maximal indique la séquence peptidique la plus probable selon le patron de fragmentation observé et selon l'information présentée dans la banque de données. Il faut également noter que seulement la position des pics et non leur abondance est considérée dans ces analyses.

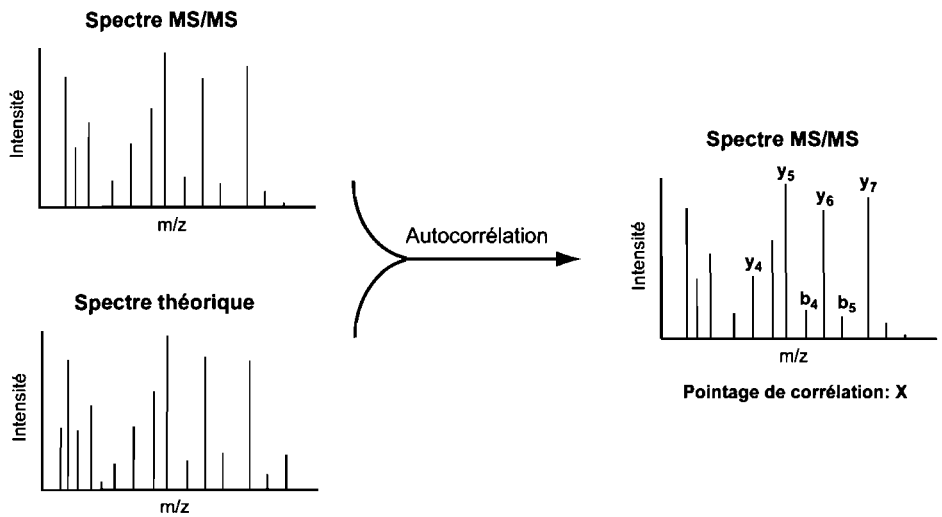

Figure 1.7: Corrélation des spectres MS/MS par la recherche dans les banques de données.

Cette méthodologie est beaucoup plus rapide et sensible pour l'identification des protéines, car seulement une fraction partielle de la séquence peptidique est nécessaire pour établir une corrélation avec un spectre théorique. Il est donc possible d'identifier une protéine présente en faible abondance à partir du spectre MS/MS d'une seule séquence peptidique partielle. Par contre, l'information présentée dans la banque de données doit être la plus complète possible et des critères stricts de validation doivent être établis pour limiter le taux de fausses identifications.

1.3.3 La quantification relative en protéomique

Le protéome de la cellule est en constant processus d'évolution en relation avec les changements induits par la stimulation de son environnement, les agents chimiques, les drogues, la croissance cellulaire ou bien par le développement de maladies. Un des grands défis en protéomique est de pouvoir quantifier ces changements qui sont observés au sein du protéome permettant ainsi d'avoir une meilleure compréhension des mécanismes biologiques qui régissent les pathologies complexes (e.g. cancer). Pour ce faire, trois principales approches furent développées par LC-MS pour faire la quantification relative des protéines à travers différentes conditions expérimentales : le marquage isotopique (IL)-MS (Figure 1.8A), le IL-MS/MS (Figure 1.8B) ainsi que l'approche sans IL (Figure 1.8C).

Dans la méthodologie par IL-MS, deux conditions sont différenciées par une comparaison de formes isotopiques « lourdes » et « légères » dans une même analyse LC-MS. L'étiquette isotopique dans chacune des conditions expérimentales peut être incorporée de façon biologique ou chimique [70]. Dans cette approche, chaque peptide se retrouve avec une forme isotopique « lourde » et « légère » dont la différence de masse entre les deux espèces est en dépendance avec le marquage isotopique effectué (e.g. ^{13}C, ^{15}N, L-leucine deutérée) et la structure du composé. Étant donné que les deux formes isotopiques sont chimiquement différentes, deux profils d'élution sont observés pour chacun des peptides dont l'abondance relative est déterminée à partir des ratios d'intensité des signaux MS des ions peptidiques « légers » et « lourds » (Figure 1.8A). Il faut prendre en compte que l'incorporation biologique est limitée aux applications utilisant des cellules en culture et que le IL chimique est exposé aux réactions secondaires potentielles.

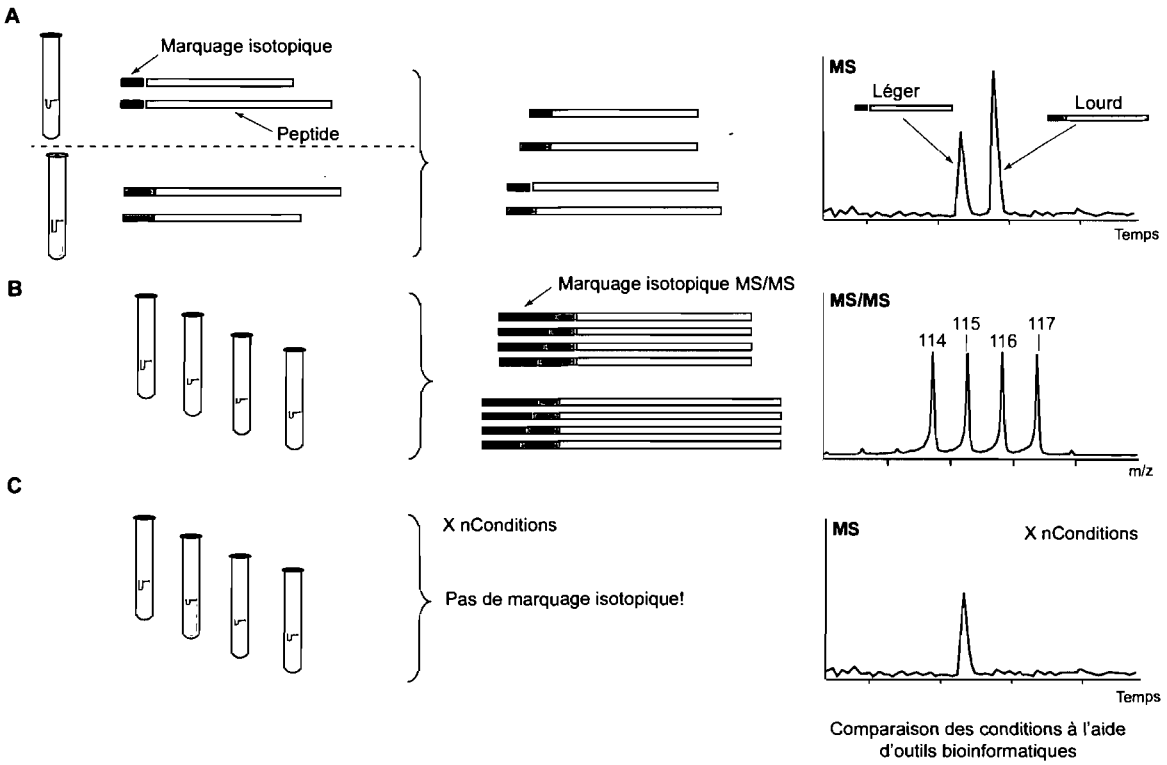

Figure 1.8: Différentes approches pour la quantification relative en MS. (A) IL-MS. Le bleu représente le IL-léger et le rouge le IL-lourd. (B) IL-MS/MS. Le bleu représente le IL-MS/MS qui diffère en masse pour chaque condition et le rouge la balance du IL-MS. (C) Sans IL. Modifiée de [45].

Une deuxième approche proposée est basée sur l'incorporation d'une étiquette isotopique MS/MS dans laquelle chaque condition contient son propre ion fragment cible (Figure 1.8B). Dans cette méthodologie, on incorpore chimiquement un groupement isobarique possédant un fragment rapporteur différent de une unité de masse pour chacune des conditions expérimentales étudiées (Figure 1.8B). Ces fragments rapporteurs sont observés lors de l'analyse MS/MS et la quantification est effectuée par comparaison des intensités des ions fragments cibles qui sont présentés dans chacun des spectres MS/MS des différents ions peptidiques [71]. Avec cette approche, il est possible de comparer jusqu'à 4 conditions expérimentales en seulement une analyse LC-MS ce qui permet d'augmenter la sensibilité MS étant donné que la masse de l'ion peptidique ciblé reste la même d'une condition à l'autre (groupement isobarique). La contrainte majeure de cette méthode réside dans le fait que la quantification peut seulement être effectuée sur des candidats dont les spectres MS/MS permettent l'identification des fragments rapporteurs et de la structure peptidique.

Finalement, la dernière approche sans IL est basée sur la comparaison de profils MS à travers différentes conditions sans aucune modification apportée à l'échantillon d'origine (Figure 1.8C). On retrouve donc un profil MS pour chacun des ions peptidiques dont les coordonnées peuvent être facilement retracées et comparées étant donné que les structures peptidiques restent inchangées. La quantification est effectuée par la comparaison des intensités MS pour chacun des ions peptidiques à travers les différentes conditions expérimentales à l'aide de logiciels spécialisés (voir section 2.4). Cette méthodologie a l'avantage de pouvoir être utilisée sur un nombre infini de conditions comparativement aux approches par IL qui sont beaucoup plus limitées. Par contre, la quantification sans IL requiert d'avoir des méthodes LC-MS reproductibles ainsi que des outils bioinformatiques appropriés pour faire une comparaison efficace des intensités à travers les conditions [30, 54, 72].

1.4 L'immunoprotéomique

La caractérisation des peptides MHC I présentés à la surface de cellules infectées ou néoplasiques est un défi d'importance pour les immunologistes pouvant avoir des répercussions majeures sur l'avancement de nouvelles thérapies contre différentes maladies. La majorité des antigènes connus actuellement ont été identifiés par différentes approches immunologiques, génétiques et protéomiques [43]. Cependant, uniquement les analyses biochimiques permettent d'évaluer quels peptides MHC I sont réellement présentés à la surface des cellules. Les premières caractérisations biochimiques du répertoire peptidique MHC I (MIP) furent effectuées à l'aide du séquençage par la dégradation d'Edman [26]. Cette méthode utilisée pour l'identification structurale fut rapidement remplacée par l'immunoprotéomique, une approche étant beaucoup plus rapide, sensible et efficace pour l'étude du répertoire MIP présenté à la surface des cellules [24, 73]. L'analyse du répertoire MIP est un défi de taille étant donné qu'on retrouve des milliers de peptides différents à la surface d'une même cellule qui sont présentés à un niveau de 1-10 000 molécules/cellule [24, 25]. Les techniques par nanoLC-MS sont actuellement les plus utilisées car elles rendent possible la séparation et l'identification de peptides MHC I présentés à des quantités de l'ordre du femtomole (1-2 fmol) au sein des matrices biologiques complexes MHC I [74, 75].

1.4.1 L'isolation des peptides du MHC I

Puisque les molécules MHC I sont des protéines membranaires et qu'elles sont présentées en un faible nombre de molécules/cellule, il est difficile de les purifier en grande quantité comme exigée pour l'analyse MS des peptides. Différentes méthodologies furent proposées pour isoler les peptides MHC I présentés à la surface des cellules. La première approche est basée sur l'extraction des molécules MHC I par lyse cellulaire en présence de détergents suivi d'une immunopurification (IP) des molécules ciblées (tableau 1.2) [76-78]. Cette technique est la plus couramment utilisée pour l'analyse des peptides MHC I en MS car elle permet d'enrichir de façon significative les extraits peptidiques en plus de pouvoir être appliquée à l'analyse de tous les modèles cellulaires (« *in vitro* » et « *in vivo* »). Par contre, cette approche est limitée pour l'étude du répertoire MIP présenté sur des cellules ou des tissus de patients parce que les rendements d'extraction sont faibles et les coûts relativement élevés (utilisation d'anticorps pour la purification).

Pour contrer ce problème, deux méthodes furent développées pour augmenter le rendement d'extraction en peptides MHC I en partant d'un nombre limité de cellules. La première approche est basée sur la transfection de cellules par l'expression de vecteurs codant la formation de molécules MHC I solubles (sMHC, tableau 1.2) [79-83]. Ces sMHC sont sécrétées dans le milieu de culture dû au manque d'un domaine fonctionnel transmembranaire. Cette particularité permet de recueillir des molécules MHC I à de multiples reprises à partir d'un nombre limité de cellules pouvant ensuite être purifiés par immunopurification. En comparaison avec l'approche par lyse cellulaire, les préparations peptidiques ne contiennent pas de contaminations provenant de détergents ou de débris cellulaires simplifiant ainsi l'analyse par nanoLC-MS [43]. Cependant, cette technique peut seulement être appliquée à des modèles cellulaires « *in vitro* » (cultures cellulaires). De plus, la transfection cellulaire peut avoir une influence sur la présentation MHC I et le répertoire MIP présenté par les sMHC ne correspond pas nécessairement à la population peptidique normalement présentée à la surface des cellules.

La seconde approche permettant d'augmenter le rendement d'extraction MHC I consiste en une élution acide douce des peptides présentés à la surface des cellules. Le milieu acide permet la dénaturation des molécules MHC I et libère la β_2-microglobuline et

les peptides MHC I dans le tampon d'élution tout en gardant la majorité des cellules viables [84]. Cette technique est efficace, rapide, peu dispendieuse et peut être appliquée à tous les modèles cellulaires. De plus, l'élution acide permet d'isoler l'ensemble du répertoire MIP présenté à la surface des cellules en une seule étape incluant les peptides ayant une plus faible affinité pour les molécules du MHC I. Par contre, cette approche a été jusqu'à présent peu utilisée pour l'analyse des peptides MHC I par nanoLC-MS étant donné que ces préparations MHC I peuvent contenir des artéfacts peptidiques dérivant d'une lyse cellulaire potentielle [85, 86]. Le tableau 1.2 fait un résumé des avantages et des désavantages de chacune des méthodologies.

Tableau 1.2: Comparaison des différentes méthodologies pour l'isolation des peptides MHC I.

	Lysats cellulaires	MHCs	Élution acide
Purification IP	oui	oui	non
Rendement	+	+++	+++
Coût	+++	+++	+
Applicabilité	*In vitro, In vivo*	*In vitro*	*In vitro, In vivo*
Contaminations possibles	Débris cellulaires, détergents & artéfacts IP	Artéfacts IP	Artéfacts (lyse cellulaire)
Débit	+	++	+++
Contraintes	Perte des peptides ayant une plus faible affinité pour le MHC I?	Transfection vs. présentation MHC I?	Contrôle de la viabilité cellulaire

Les symboles +, ++ et +++ indiquent faible, modéré ou élevé, et très élevé, respectivement. IP, immunopurification; MHCs, molécules MHC solubles

1.4.2 Identification et quantification du répertoire MIP

Les techniques par nanoLC-MS sont présentement les plus répandues pour l'analyse du répertoire MIP. Pour résoudre la grande complexité associée aux préparations MHC I, les extraits sont dans la majorité du temps fractionnés par chromatographie à résine échangeuse d'ions ou à polarité de phase inversée (RP) préalablement aux analyses par

nanoLC-MS [76, 87]. Étant donné le défi face à la caractérisation des peptides MHC I, très peu d'études se sont concentrées sur l'analyse différentielle du répertoire MIP [30, 77, 83]. Certaines approches utilisées pour la quantification relative en protéomique ont été appliquées à l'analyse à haut débit des peptides MHC I apportant ainsi une information cruciale à la meilleure compréhension des mécanismes contrôlant la présentation antigénique ainsi que l'influence de la transformation néoplasique sur la composition du répertoire MIP.

Parmi ces techniques, on compte le IL métabolique [83, 87], le IL chimique [76, 77] ainsi que celle sans marquage. L'approche sans marquage fut proposée pour la première fois pour l'analyse différentielle du répertoire MIP dans le cadre des travaux de recherche décrits dans cette thèse [30]. Le tableau 1.3 compare les avantages et les inconvénients de chacune des méthodologies qui sont expliquées en plus de détails dans la section 1.3.3.

Tableau 1.3: Comparaison des différentes méthodologies utilisées pour la quantification relative du répertoire MIP.

	IL métabolique	IL chimique	Sans marquage
Applicabilité	In vitro	In vitro, In vivo	In vitro, In vivo
Complexité des échantillons	+++	+++	++
Coût : préparation	+++	+++	+
Débit : préparation	+	++	+++
Débit : analyses MS	+++	+++	++
Contraintes	Limité au niveau des applications biologiques	Réactions secondaires possibles	Besoin de méthodes LC-MS reproductibles & d'outils bioinformatiques efficaces

Les symboles +, ++ et +++ indiquent faible, modéré ou élevé, et très élevé, respectivement. IL, marquage isotopique.

En ce qui concerne la quantification relative du répertoire MIP, les techniques utilisant le marquage isotopique sont plus limitées étant donné qu'elles augmentent la complexité des échantillons (formes « légères » et « lourdes »). Les coûts reliés à ces

techniques de IL sont d'ailleurs relativement élevés dans le cadre de l'étude du répertoire MIP en relation avec la quantité de matériel nécessaire pour l'analyse des peptides MHC I par nanoLC-MS. Le IL métabolique présente encore plus de contraintes pour l'analyse différentielle du répertoire MIP, car cette approche peut seulement être utilisée pour des modèles cellulaires « *in vitro* » en plus d'être limitée à certaines classes de molécules MHC I (acides aminés incorporés doivent être présentés au niveau du motif).

1.5 Objectifs de la recherche

L'objectif principal de ces travaux de recherche était de développer une nouvelle méthodologie analytique pour donner une définition moléculaire précise du répertoire MIP présenté par des cellules normales et néoplasiques. L'évaluation de différents systèmes chromatographiques fut effectuée pour tenter de maximiser la sensibilité et la sélectivité de la séparation chromatographique. Dans le but de répondre à cette objectif, les performances analytiques des colonnes capillaires (nanoLC-MS) furent comparées à celles d'un nouveau système microfluidique (nanoLC-chip-MS, Agilent Technologies). Cependant, étant donné que les performances des deux systèmes étaient comparables et que l'instrumentation MS pouvant être couplée à la puce nanoLC était plus limitée, l'approche utilisant les colonnes capillaires fut préférée pour l'étude du répertoire MIP.

Pour résoudre la complexité associée aux populations peptidiques MHC I, une approche protéomique combinant différentes stratégies bioinformatiques à des techniques analytiques de séparation par nanoLC-MS fut donc proposée pour identifier des candidats potentiels montrant des différences d'expression significatives entre des cellules normales et leucémiques. Un des buts de cette étude comparative était dans un premier temps d'identifier des peptides cibles surexprimés ou uniquement exprimés au niveau des cellules leucémiques pouvant être potentiellement utilisés pour le développement de vaccins contre le cancer du sang. L'avancement de la recherche au niveau de ces nouvelles thérapies est présentement limité dû au manque de connaissances en ce qui a attrait à la diversité de la population allogénique présentée à la surface des cellules néoplasiques.

Dans un deuxième temps, l'objectif des études présentées dans cette thèse était de répondre à certaines questions biologiques pour permettre une meilleure compréhension des mécanismes régulant la genèse et la composition du répertoire MIP. Dans ce contexte biologique, cinq questions spécifiques furent adressées : 1) Est-ce que le répertoire MIP reflète la composition du transcriptome? 2) Est-ce que certaines familles de gènes sont plus représentées dans le répertoire MIP? 3) Quel est l'impact de la transformation néoplasique sur le répertoire MIP? 4) Est-ce que la voie de signalisation mTOR peut avoir une influence majeure sur la composition du répertoire MIP? 5) Est-ce qu'une approche analytique par

spectrométrie de masse à haut débit peut permettre l'identification d'antigènes tumorales immunogéniques?

1.6 Organisation de la thèse

Comme introduction à la démarche expérimentale, le **deuxième chapitre** de cette thèse présente un survol de la méthodologie ainsi que des techniques analytiques qui furent employées dans le cadre de ces travaux de recherche. Les chapitres suivants (3, 4 et 5) sont ensuite présentés sous forme d'article scientifique.

Le **chapitre 3** est consacré à l'évaluation des performances chromatographiques d'un nouveau système microfluidique conçu par Agilent Technologies [53]. Les performances analytiques des colonnes capillaires (nanoLC-MS) furent comparées à celles du nouveau système microfluidique (nanoLC-chip-MS). Les limitations et les avantages des deux systèmes en terme de capacité chromatographique, de reproductibilité et de sensibilité furent comparés à partir d'analyses effectuées sur des échantillons protéiques complexes. Ces études ont permis d'établir que les performances des deux systèmes étudiés sont comparables en terme de limite de détection ainsi qu'au niveau de la reproductibilité des intensités. Les travaux présentés dans ce chapitre sont distincts de ceux des chapitres 4 et 5. La puce nanoLC ne fut pas utilisée pour les autres travaux présentés dans cette thèse.

Le **chapitre 4** présente une nouvelle méthodologie analytique combinant différentes stratégies bioinformatiques à des techniques analytiques de séparation par 2D-nanoLC-MS afin d'obtenir une définition moléculaire précise du répertoire MIP de cellules normales et leucémiques. Cette étude comparative a permis d'identifier des peptides surexprimés ou uniquement exprimés au niveau des cellules leucémiques pouvant être potentiellement utilisés dans le développement de vaccins contre le cancer du sang. Les essais d'immunisation effectués avec deux candidats ont d'ailleurs montré que des changements en expression d'un facteur de 10 à $\geq$ 85 fois pouvaient stimuler une réponse cytotoxique spécifique. Cette étude propose aussi des concepts biologiques pouvant permettre de mieux comprendre les mécanismes qui régissent la genèse et la composition du répertoire MIP.

Le **chapitre 5** se concentre sur l'étude de l'influence de la voie de signalisation mTOR sur la genèse et la composition du répertoire MIP. La voie oncogénique mTOR est connue pour jouer un rôle déterminant dans la régulation de la synthèse et la dégradation

des protéines [9], deux processus dynamiques ayant une influence significative sur la présentation antigénique [40]. La méthodologie analytique proposée dans le chapitre 4 fut employée pour étudier la cinétique de présentation MHC I suite à l'inhibition de mTOR dans des cellules néoplasiques. Les résultats présentés indiquent que l'inhibition de mTOR augmente la présentation antigénique à la surface de cellules néoplasiques de façon significative. Ces données suggèrent que la dérégulation de la voie de signalisation mTOR lors de la transformation néoplasique peut influencer de façon importante la genèse et la composition globale du répertoire MIP.

Finalement, le **chapitre 6** présente une synthèse de l'ensemble des résultats présentés dans le cadre de ces travaux de recherche ainsi que quelques perspectives en lien avec les travaux en cours.

1.7 Références

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *Biologie Moléculaire de la Cellule, Édition française*. 4 ed. 2004, Flammarion Médecine-Sciences, Paris, France: p. 1313-1421.

2. *Statistiques canadiennes sur le cancer 2008*. Institut national du cancer du Canada. 2008 [Page consultée le 14 Octobre 2008]; Disponible à partir de: http://www.ncic.cancer.ca/sitecore/content/NCIC/Cancer%20control/~/media/NCIC/Files%20list/Liste%20de%20fichiers/PDF/Canadian%20Cancer%20Statistics%20-%202008%20-%20FR%20-%20pdf_195993224.ashx.

3. Zitvogel, L., Apetoh, L., Ghiringhelli, F., and Kroemer, G., *Immunological aspects of cancer chemotherapy*. Nat Rev Immunol, 2008. **8**(1): p. 59-73.

4. Frohling, S. and Dohner, H., *Chromosomal abnormalities in cancer*. N Engl J Med, 2008. **359**(7): p. 722-34.

5. Nambiar, M., Kari, V., and Raghavan, S.C., *Chromosomal translocations in cancer*. Biochim Biophys Acta, 2008. **1786**(2): p. 139-52.

6. Kroemer, G. and Pouyssegur, J., *Tumor cell metabolism: cancer's Achilles' heel*. Cancer Cell, 2008. **13**(6): p. 472-82.

7. Zitvogel, L., Tesniere, A., and Kroemer, G., *Cancer despite immunosurveillance: immunoselection and immunosubversion*. Nat Rev Immunol, 2006. **6**(10): p. 715-27.

8. Chiang, G.G. and Abraham, R.T., *Targeting the mTOR signaling network in cancer*. Trends Mol Med, 2007. **13**(10): p. 433-42.

9. Guertin, D.A. and Sabatini, D.M., *Defining the role of mTOR in cancer*. Cancer Cell, 2007. **12**(1): p. 9-22.

10. Vivanco, I. and Sawyers, C.L., *The phosphatidylinositol 3-Kinase AKT pathway in human cancer*. Nat Rev Cancer, 2002. **2**(7): p. 489-501.

11. Yang, Q. and Guan, K.L., *Expanding mTOR signaling*. Cell Res, 2007. **17**(8): p. 666-81.

12. Fingar, D.C. and Blenis, J., *Target of rapamycin (TOR): an integrator of nutrient and growth factor signals and coordinator of cell growth and cell cycle progression*. Oncogene, 2004. **23**(18): p. 3151-71.

13. Levine, B. and Kroemer, G., *Autophagy in the pathogenesis of disease.* Cell, 2008. **132**(1): p. 27-42.

14. Holland, E.C., Sonenberg, N., Pandolfi, P.P., and Thomas, G., *Signaling control of mRNA translation in cancer pathogenesis.* Oncogene, 2004. **23**(18): p. 3138-44.

15. Tsang, C.K., Qi, H., Liu, L.F., and Zheng, X.F., *Targeting mammalian target of rapamycin (mTOR) for health and diseases.* Drug Discov Today, 2007. **12**(3-4): p. 112-24.

16. Faivre, S., Kroemer, G., and Raymond, E., *Current development of mTOR inhibitors as anticancer agents.* Nat Rev Drug Discov, 2006. **5**(8): p. 671-88.

17. Altman, J.K. and Platanias, L.C., *Exploiting the mammalian target of rapamycin pathway in hematologic malignancies.* Curr Opin Hematol, 2008. **15**(2): p. 88-94.

18. Rosenberg, S.A., *Progress in human tumour immunology and immunotherapy.* Nature, 2001. **411**(6835): p. 380-4.

19. Begley, J. and Ribas, A., *Targeted therapies to improve tumor immunotherapy.* Clin Cancer Res, 2008. **14**(14): p. 4385-91.

20. Purcell, A.W., McCluskey, J., and Rossjohn, J., *More than one reason to rethink the use of peptides in vaccine design.* Nat Rev Drug Discov, 2007. **6**(5): p. 404-14.

21. Roitt, I., Brostoff, J., and Male, D., *Immunologie, édition française.* 3 ed. 2002, De Boeck & Larcier, Bruxelles, Belgique: 1-118.

22. Klein, J. and Sato, A., *The HLA system. First of two parts.* N Engl J Med, 2000. **343**(10): p. 702-9.

23. Perreault, C. and Brochu, S., *Adoptive cancer immunotherapy: discovering the best targets.* J Mol Med, 2002. **80**(4): p. 212-8.

24. Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., and Engelhard, V.H., *Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry.* Science, 1992. **255**(5049): p. 1261-3.

25. Rammensee, H.G., Falk, K., and Rotzschke, O., *Peptides naturally presented by MHC class I molecules.* Annu Rev Immunol, 1993. **11**: p. 213-44.

26. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., and Rammensee, H.G., *Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules.* Nature, 1991. **351**(6324): p. 290-6.

27. Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D.C., Sidney, J., Bui, H.H., Grey, H., and Sette, A., *A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus.* Nat Biotechnol, 2006. **24**(7): p. 817-9.

28. Peters, B. and Sette, A., *Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method.* BMC Bioinformatics, 2005. **6**: p. 132.

29. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S., *SYFPEITHI: database for MHC ligands and peptide motifs.* Immunogenetics, 1999. **50**(3-4): p. 213-9.

30. Fortier, M.H., Caron, E., Hardy, M.P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P., *The MHC class I peptide repertoire is molded by the transcriptome.* J Exp Med, 2008. **205**(3): p. 595-610.

31. Kloetzel, P.M., *Antigen processing by the proteasome.* Nat Rev Mol Cell Biol, 2001. **2**(3): p. 179-87.

32. Hammer, G.E., Kanaseki, T., and Shastri, N., *The final touches make perfect the peptide-MHC class I repertoire.* Immunity, 2007. **26**(4): p. 397-406.

33. Cresswell, P., Ackerman, A.L., Giodini, A., Peaper, D.R., and Wearsch, P.A., *Mechanisms of MHC class I-restricted antigen processing and cross-presentation.* Immunol Rev, 2005. **207**: p. 145-57.

34. Elliott, T. and Williams, A., *The optimization of peptide cargo bound to MHC class I molecules by the peptide-loading complex.* Immunol Rev, 2005. **207**: p. 89-99.

35. Jensen, P.E., *Recent advances in antigen processing and presentation.* Nat Immunol, 2007. **8**(10): p. 1041-8.

36. Vyas, J.M., Van der Veen, A.G., and Ploegh, H.L., *The known unknowns of antigen processing and presentation.* Nat Rev Immunol, 2008. **8**(8): p. 607-18.

37. Yewdell, J.W., Reits, E., and Neefjes, J., *Making sense of mass destruction: quantitating MHC class I antigen presentation.* Nat Rev Immunol, 2003. **3**(12): p. 952-61.

38. Yewdell, J.W. and Nicchitta, C.V., *The DRiP hypothesis decennial: support, controversy, refinement and extension.* Trends Immunol, 2006. **27**(8): p. 368-73.

39. Schubert, U., Anton, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., and Bennink, J.R., *Rapid degradation of a large fraction of newly synthesized proteins by proteasomes.* Nature, 2000. **404**(6779): p. 770-4.

40. Yewdell, J.W., *Plumbing the sources of endogenous MHC class I peptide ligands.* Curr Opin Immunol, 2007. **19**(1): p. 79-86.

41. Reits, E.A., Vos, J.C., Gromme, M., and Neefjes, J., *The major substrates for TAP in vivo are derived from newly synthesized proteins.* Nature, 2000. **404**(6779): p. 774-8.

42. Eisenlohr, L.C., Huang, L., and Golovina, T.N., *Rethinking peptide supply to MHC class I molecules.* Nat Rev Immunol, 2007. **7**(5): p. 403-10.

43. Admon, A., Barnea, E., and Ziv, T., *Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides.* Mol Cell Proteomics, 2003. **2**(6): p. 388-98.

44. Aebersold, R. and Goodlett, D.R., *Mass spectrometry in proteomics.* Chem Rev, 2001. **101**(2): p. 269-95.

45. Domon, B. and Aebersold, R., *Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-7.

46. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

47. Karas, M. and Hillenkamp, F., *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal Chem, 1988. **60**(20): p. 2299-301.

48. Liebler, D.C., *Introduction to Proteomics: Tools for the New Biology.* 2002, Humana Press, Totowa, NY: p. 3-13.

49. Liebler, D.C., *Introduction to Proteomics: Tools for the New Biology.* 2002, Humana Press, Totowa, NY: p. 31-48.

50. Gooding, K.M. and Regnier, F.E., *HPLC of Biological Macromolecules.* 2 ed. Chromatographic Science Series. Vol. 87. 2002, Marcel Dekker, New York, NY: p. 689-735.

51. Bonner, P.L., Lill, J.R., Hill, S., Creaser, C.S., and Rees, R.C., *Electrospray mass spectrometry for the identification of MHC class I-associated peptides expressed on cancer cells.* J Immunol Methods, 2002. **262**(1-2): p. 5-19.

52. Niessen, W.M.A., *Liquid Chromatography-Mass Spectrometry.* 3 ed. Chromatographic Science Series. Vol. 97. 2006, CRC, Taylor and Francis Group, Boca Raton, FL: p. 463-492.

53. Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., and van de Goor, T., *Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip.* Anal Chem, 2005. **77**(2): p. 527-33.

54. Fortier, M.H., Bonneil, E., Goodley, P., and Thibault, P., *Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs.* Anal Chem, 2005. **77**(6): p. 1631-40.

55. Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., and Mann, M., *Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry.* Nature, 2002. **419**(6906): p. 537-42.

56. Aebersold, R. and Mann, M., *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.

57. Steen, H. and Mann, M., *The ABC's (and XYZ's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.

58. Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., 3rd, *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.

59. Washburn, M.P., Wolters, D., and Yates, J.R., 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.

60. Watson, J.T. and Sparkman, O.D., *Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation.* 4 ed. 2007, John Wiley & Sons, Chichester, England: p. 1-27.

61. Cech, N.B. and Enke, C.G., *Practical implications of some recent studies in electrospray ionization fundamentals.* Mass Spectrom Rev, 2001. **20**(6): p. 362-87.

62. Cole, R.B., *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications.* 1997, John Wiley & Sons, New York, NY: p. 6-63.

63. Watson, J.T. and Sparkman, O.D., *Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation.* 4 ed. 2007, John Wiley & Sons, Chichester, England: p. 497-517.

64. Wilm, M. and Mann, M., *Analytical properties of the nanoelectrospray ion source.* Anal Chem, 1996. **68**(1): p. 1-8.

65. Scigelova, M. and Makarov, A., *Orbitrap mass analyzer--overview and applications in proteomics.* Proteomics, 2006. **6 (Suppl 2)**: p. 16-21.

66. Kinter, M. and Sherman, N.E., *Protein Sequencing and Identification Using Tandem Mass Spectrometry.* 2000, John Wiley & Sons, New York, NY: p. 64-80.

67. Snyder, A.P., *Interpreting Protein Mass Spectra: A Comprehensive Resource.* 2000, Oxford University Press, New York, NY: p. 155-164.

68. Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., and Lajoie, G., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.

69. Liebler, D.C., *Introduction to Proteomics: Tools for the New Biology.* 2002, Humana Press, Totowa, NY: p. 99-108.

70. Ong, S.E., Foster, L.J., and Mann, M., *Mass spectrometric-based approaches in quantitative proteomics.* Methods, 2003. **29**(2): p. 124-30.

71. Unwin, R.D., Smith, D.L., Blinco, D., Wilson, C.L., Miller, C.J., Evans, C.A., Jaworska, E., Baldwin, S.A., Barnes, K., Pierce, A., Spooncer, E., and Whetton, A.D., *Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells.* Blood, 2006. **107**(12): p. 4687-94.

72. Marcantonio, M., Trost, M., Courcelles, M., Desjardins, M., and Thibault, P., *Combined enzymatic and data mining approaches for comprehensive phosphoproteome analyses: application to cell signaling events of interferon-gamma-stimulated macrophages.* Mol Cell Proteomics, 2008. **7**(4): p. 645-60.

73. Purcell, A.W. and Gorman, J.J., *Immunoproteomics: Mass spectrometry-based methods to study the targets of the immune response.* Mol Cell Proteomics, 2004. **3**(3): p. 193-208.

74. Lemmel, C. and Stevanovic, S., *The use of HPLC-MS in T-cell epitope identification.* Methods, 2003. **29**(3): p. 248-59.

75. Purcell, A.W., *Isolation and characterization of naturally processed MHC-bound peptides from the surface of antigen-presenting cells.* Methods Mol Biol, 2004. **251**: p. 291-306.

76. Lemmel, C., Weik, S., Eberle, U., Dengjel, J., Kratt, T., Becker, H.D., Rammensee, H.G., and Stevanovic, S., *Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling.* Nat Biotechnol, 2004. **22**(4): p. 450-4.

77. Weinzierl, A.O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.G., and Stevanovic, S., *Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface.* Mol Cell Proteomics, 2007. **6**(1): p. 102-13.

78. Zarling, A.L., Polefrone, J.M., Evans, A.M., Mikesh, L.M., Shabanowitz, J., Lewis, S.T., Engelhard, V.H., and Hunt, D.F., *Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy.* Proc Natl Acad Sci U S A, 2006. **103**(40): p. 14889-94.

79. Barnea, E., Beer, I., Patoka, R., Ziv, T., Kessler, O., Tzehoval, E., Eisenbach, L., Zavazava, N., and Admon, A., *Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens.* Eur J Immunol, 2002. **32**(1): p. 213-22.

80. Buchsbaum, S., Barnea, E., Dassau, L., Beer, I., Milner, E., and Admon, A., *Large-scale analysis of HLA peptides presented by HLA-Cw4.* Immunogenetics, 2003. **55**(3): p. 172-6.

81. Hickman, H.D., Luis, A.D., Bardet, W., Buchli, R., Battson, C.L., Shearer, M.H., Jackson, K.W., Kennedy, R.C., and Hildebrand, W.H., *Cutting edge: class I presentation of host peptides following HIV infection.* J Immunol, 2003. **171**(1): p. 22-6.

82. Hickman, H.D., Luis, A.D., Buchli, R., Few, S.R., Sathiamurthy, M., VanGundy, R.S., Giberson, C.F., and Hildebrand, W.H., *Toward a definition of self: proteomic evaluation of the class I peptide repertoire.* J Immunol, 2004. **172**(5): p. 2944-52.

83. Milner, E., Barnea, E., Beer, I., and Admon, A., *The turnover kinetics of major histocompatibility complex peptides of human cancer cells.* Mol Cell Proteomics, 2006. **5**(2): p. 357-65.

84. Storkus, W.J., Zeh, H.J., 3rd, Salter, R.D., and Lotze, M.T., *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.

85. Herr, W., Ranieri, E., Gambotto, A., Kierstead, L.S., Amoscato, A.A., Gesualdo, L., and Storkus, W.J., *Identification of naturally processed and HLA-presented Epstein-Barr virus peptides recognized by CD4(+) or CD8(+) T lymphocytes from human blood.* Proc Natl Acad Sci U S A, 1999. **96**(21): p. 12033-8.

86. McBride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A., Bergeron, J., and Perreault, C., *The model B6(dom1) minor histocompatibility antigen is encoded by a mouse homolog of the yeast STT3 gene.* Immunogenetics, 2002. **54**(8): p. 562-9.

87. Meiring, H.D., Soethout, E.C., Poelen, M.C., Mooibroek, D., Hoogerbrugge, R., Timmermans, H., Boog, C.J., Heck, A.J., de Jong, A.P., and van Els, C.A., *Stable isotope tagging of epitopes: a highly selective strategy for the identification of major histocompatibility complex class I-associated peptides induced upon viral infection.* Mol Cell Proteomics, 2006. **5**(5): p. 902-13.

2. Méthodologie

2.1 L'étude du répertoire MIP de cellules normales et néoplasiques

Au moment de l'initiation du projet, aucune étude comparative à grande échelle n'avait été effectuée entre le répertoire MIP de cellules normales et néoplasiques. Il était connu par les immunologistes que la transformation néoplasique affectait la composition du répertoire MIP mais aucune donnée scientifique indiquait l'ampleur des différences. Les connaissances fragmentaires sur le sujet limitaient d'ailleurs le développement de nouvelles thérapies immunologiques pour le traitement du cancer. L'objectif principal du projet de recherche était donc de développer une méthodologie analytique à haut débit pour donner une définition moléculaire précise du répertoire MIP présenté par des cellules normales et néoplasiques. Cette approche analytique pourrait ainsi grandement contribuer à fournir une information cruciale pour améliorer notre compréhension sur la genèse et la composition du répertoire MIP.

2.2 L'isolation du répertoire MIP

Dans le cadre de ces travaux de recherche, l'approche basée sur l'élution acide des peptides présentés à la surface des cellules par dénaturation des molécules MHC I fut utilisée pour isoler le répertoire MIP de cellules normales et leucémiques (voir section 1.4.1 pour plus de détails). Les avantages face à l'utilisation de l'élution acide par rapport aux autres approches ont justifié le choix de cette technique pour les études décrites dans la présente thèse. Nous cherchions une technique rapide et efficace pour isoler l'ensemble du répertoire MIP présenté à la surface des cellules. L'élution acide est présentement la seule approche permettant d'isoler le répertoire MIP qui est réellement présenté à la surface de cellules fraîchement extraites. Cette technique a d'ailleurs l'avantage de pouvoir être utilisée sur tous les modèles cellulaires (« *in vivo* » et « *in vitro* »). L'élution acide permet d'obtenir un rendement ~10 fois supérieur aux approches utilisant les immunopurifications ce qui la rend accessible à un plus grand nombre d'applications « *in vivo* » [1]. La seule contrainte majeure est reliée au fait qu'il faut avoir un bon contrôle sur la viabilité cellulaire pour limiter la présence de peptides artéfacts dans les préparations peptidiques MHC I. Ce problème peut facilement être résolu dans le cadre d'analyses « *in vitro* » par l'utilisation d'un contrôle négatif β_2-microglobulin KO (β_2m^-) comme expliqué plus en détails dans le

chapitre 4 [2]. Brièvement, la diminution significative en abondance d'un peptide X dans la condition β_2m^- permet de confirmer l'affinité de ce peptide X pour le MHC I étant donné que l'expression MHC I à la surface est pratiquement inexistante dans la condition β_2m^-. Dans le cadre d'analyses « *in vivo* », l'utilisation d'un contrôle β_2m^- étant impossible, il fallait trouver une méthodologie efficace pour confirmer les structures MHC I spécifiques au sein des extraits peptidiques. Les analyses présentées dans le chapitre 4 ont montré que l'application de critères de sélection stricts (pointages d'affinité élevés) établis à l'aide des algorithmes de prédiction permettait d'obtenir un taux de confiance très élevé (98%) par rapport à l'identification et à la confirmation des structures MHC I spécifiques.

2.3 Méthodes de séparation par nanoLC-MS

La grande complexité face à la caractérisation des préparations MHC I demande l'utilisation d'instrumentation analytique à la fine pointe de la technologie (section 1.2.1, 1.3.1 et 1.4). Les techniques de séparation par nanoLC-MS sont actuellement les plus appropriées pour répondre aux critères de sensibilité et de résolution nécessaires à l'analyse du répertoire MIP. La prochaine section résume les différentes techniques par nanoLC-MS qui furent utilisées dans le cadre de ces travaux de recherche.

2.3.1 Système microfluidique

Dans le cadre du chapitre 3, les performances chromatographiques sont présentées pour un nouveau système microfluidique conçu par Agilent Technologies [3]. Les performances analytiques des colonnes capillaires traditionnellement utilisées en protéomique (nanoLC-MS) furent comparées à celles du nouveau système microfluidique (nanoLC-chip-MS) pour l'analyse de mélanges protéiques complexes. L'objectif principal en lien avec le projet de recherche était d'évaluer le potentiel analytique (e.g. sensibilité, reproductibilité, etc...) de ce système microfluidique pour l'analyse du répertoire MIP.

Pour des raisons pratiques, l'utilisation des systèmes microfluidiques pour la séparation des peptides fut historiquement plus développée pour des systèmes basés sur le principe de l'électrophorèse capillaire (CE) [4] et de l'électrochromatographie capillaire

[5]. La puce nanoLC conçue par Agilent Technologies est le premier système intégré combinant les différentes composantes utilisées en nanoLC-MS sur un même module compact [3]. Cette configuration permet de réduire les volumes morts par la diminution des lignes de transfert et des connexions se traduisant par une amélioration des performances chromatographiques (e.g. capacité chromatographique, largeur et symétrie des pics, sensibilité, etc...). De plus, la miniaturisation du canal de séparation augmente la sensibilité de l'analyse en permettant une meilleure concentration des bandes d'analytes [6]. La figure 2.1 montre les différentes composantes de la puce nanoLC utilisée dans le cadre des travaux du chapitre 3. Les détails par rapport à la fabrication de la puce sont décrits dans Yin et *al.* [3].

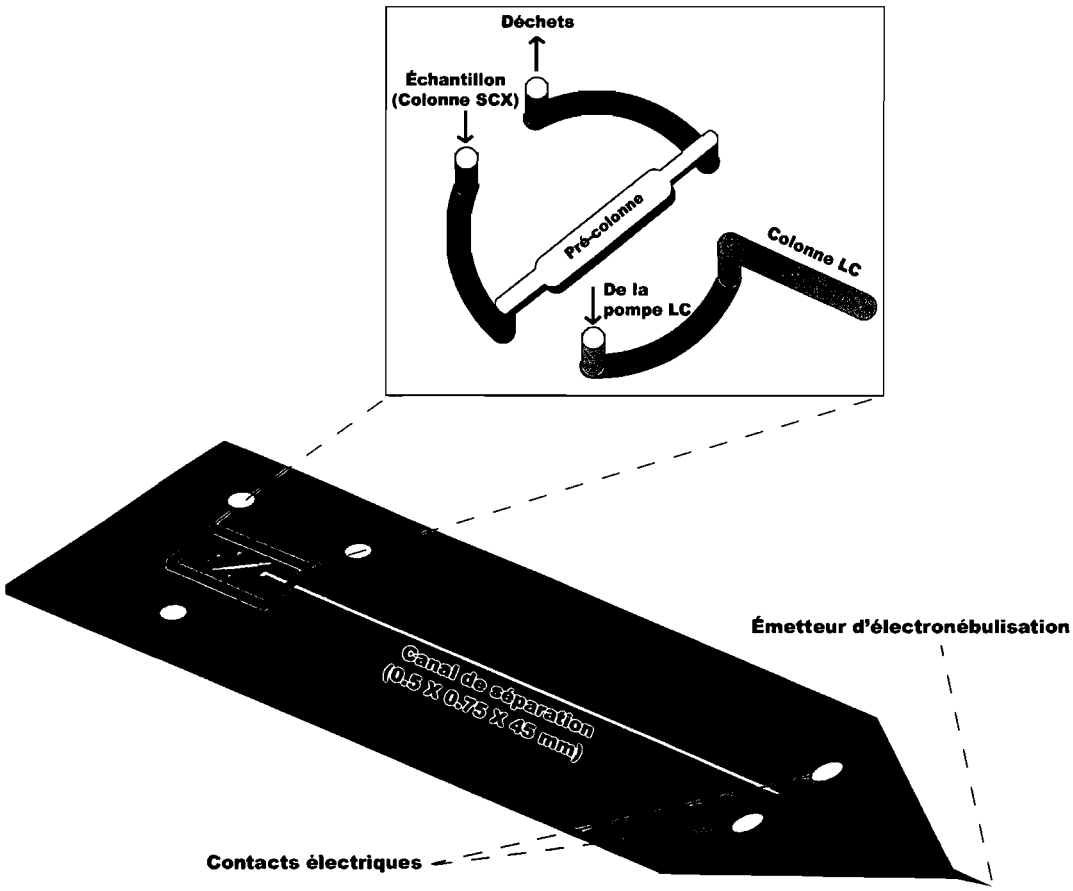

Figure 2.1: Représentation schématique de la puce nanoLC. Modifiée de [7].

Brièvement, la puce nanoLC est un système microfluidique composé d'une pré-colonne, d'une colonne chromatographique (canal de séparation) ainsi que d'un émetteur

d'électronébulisation sur un même module compact (Figure 2.1). La configuration standard de la puce nanoLC comporte une pré-colonne avec un volume interne de 40 nL, un émetteur avec un diamètre interne de 15 µm et une colonne chromatographique ayant une section transversale de 75 µm x 50 µm et une longueur de 45 mm (vol = 169 nL). La puce nanoLC est couplée au spectromètre de masse par l'intermédiaire d'une interface nanoESI conçue spécifiquement pour les débits utilisés (~ 100-600 nL/min, chapitre 3) [3, 8]. La configuration spécifique de l'interface permet d'établir le lien avec le système nanoLC-MS et la puce nanoLC par l'intermédiaire d'une micro valve qui contrôle la direction de l'éluant lors du chargement de l'échantillon et de la séparation analytique.

2.3.1.1 Interface et système nanoLC-chip-MS

La puce nanoLC fut utilisée uniquement dans le cadre des travaux présentés dans le chapitre 3. Brièvement, la puce nanoLC est positionnée entre le stator et le rotor d'une valve à deux positions Rheodyne HPLC [3]. La puce nanoLC est installée sur un support contenant le rotor de la valve et est maintenue en place par des ports d'alignement. Le stator de la valve est intégré à un mécanisme de verrouillage permettant de maintenir de façon hermétique et d'aligner correctement les ports d'entrée de la puce nanoLC avec les ports d'entrée et les canaux de la micro valve d'injection. Lors des analyses MS, l'émetteur d'électronébulisation est placé perpendiculairement au cône d'échantillonnage ainsi qu'au spectromètre de masse. Le contrôle de la micro valve d'injection et le positionnement des ports est comparable au système traditionnel nanoLC-MS. Étant donné que l'interface nanoLC-chip-MS pouvait seulement être couplée à des instruments conçus par Agilent Technologies, les analyses MS ont été effectuées sur un instrument MSD TOF (Agilent Technologies) et les analyses MS/MS sur un MSD Ion Trap (Agilent Technologies). Ces instruments ont seulement été utilisés pour les travaux présentés dans le chapitre 3. Leur principe de fonctionnement est analogue aux descriptions données pour les trappes ioniques (IT) et les spectromètres de masse temps d'envol (TOF) [9].

2.3.2 Chromatographie bidimensionnelle

Dans le cadre des travaux présentés dans les chapitres 3 à 5, la chromatographie bidimensionnelle fut utilisée pour résoudre la complexité associée aux préparations MHC I

et aux mélanges protéiques plasmiques (section 1.3.1). Une séparation orthogonale des peptides fut effectuée par combinaison de la chromatographie à échange de cations (SCX) et de la chromatographie en phase inversée (RP). La première dimension permettait donc une séparation des peptides par rapport à leur charge nette positive et la seconde selon leur hydrophobicité respective. Cette complémentarité chromatographique fut utilisée pour obtenir une meilleure séparation des populations peptidiques permettant ainsi d'augmenter notre capacité à détecter des peptides présentés en faible abondance au sein des matrices complexes étudiées. Cette technique fut aussi employée pour augmenter notre capacité de chargement en configuration nanoLC. Les deux approches qui furent utilisées dans le cadre de ces travaux de recherche sont présentées dans les sections suivantes.

2.3.2.1 Par fractionnement

Le fractionnement par chromatographie d'échange de cations préalablement à la séparation par nanoLC-MS fut utilisé pour les expériences effectuées dans le chapitre 5. Les extraits peptidiques MHC I furent séparés sur une colonne SCX à l'aide d'un gradient linéaire de formate d'ammonium et collectés en différentes fractions lesquelles furent analysées ultérieurement par nanoLC-MS. Cette approche permet une meilleure résolution chromatographique pour la première dimension de séparation et rend possible l'injection d'une plus grande quantité de matériel. Par contre, les étapes supplémentaires de manipulation des extraits augmentent les pertes potentielles de matériel rendant cette technique moins intéressante pour l'analyse par exemple d'échantillon « *in vivo* » comme présenté dans le chapitre 4.

2.3.2.2 Système en ligne

Des techniques de séparation bidimensionnelle en ligne avec les systèmes nanoLC-MS furent développées pour limiter la perte de matériel reliée au fractionnement [10, 11]. Cette méthodologie fut employée dans le cadre des travaux présentés dans les chapitres 3 et 4. Typiquement, les peptides sont chargés dans un premier temps sur une colonne SCX, reliée en ligne avec le système nanoLC-MS entre l'injecteur et la micro valve d'injection (Figure 2.2). Pour la puce nanoLC, le couplage de la colonne SCX est fait entre l'injecteur et la micro valve de l'interface nanoLC-chip-MS pour obtenir une

configuration similaire au système traditionnel nanoLC-MS (Figure 2.1). Les peptides sont retenus sur la colonne SCX en lien avec leur caractère cationique. Ceux-ci sont par la suite progressivement élués de la phase stationnaire SCX par un gradient en pallier utilisant différentes concentrations d'acétate d'ammonium.

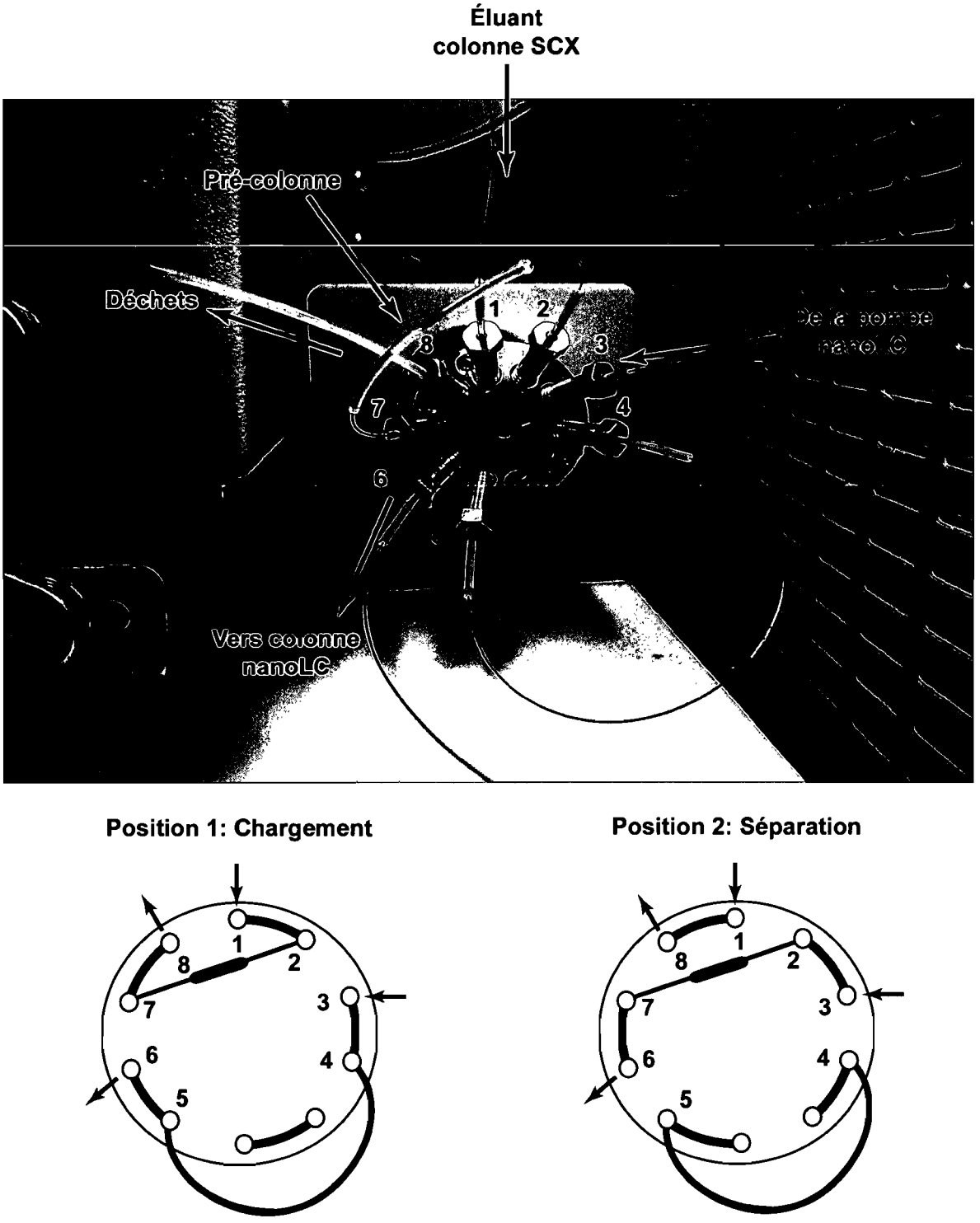

Figure 2.2: Élution des peptides en mode 2D-nanoLC-MS (système en ligne).

On élue donc une première population peptidique à une concentration X de sel qui est chargée sur la pré-colonne C_{18} (Position 1; Figure 2.2). Cette population peptidique est ensuite éluée sur la colonne C_{18} et analysée par nanoLC-MS (Position 2; Figure 2.2). Ce cycle se répète pour des concentrations croissantes de sel de façon synchronisée jusqu'à ce que tous les peptides soient élués de la colonne SCX (5 à 8 fractions). Cette approche est des plus appropriées pour l'analyse de mélanges complexes, car elle permet de résoudre la complexité tout en limitant la quantité de matériel nécessaire. Par contre, cette technique en ligne offre une résolution inférieure pour la première dimension de chromatographie et requiert plus de temps d'utilisation sur le système nanoLC-MS.

2.3.3 Spectromètre de masse LTQ-Orbitrap

Dans le cadre des travaux présentés dans les chapitres 4 et 5, un spectromètre de masse de type LTQ-Orbitrap (Thermo Fisher Scientific) fut utilisé pour faciliter l'analyse du répertoire MIP. L'hybride trappe linéaire quadrupôle (LTQ)-Orbitrap est un nouveau type d'analyseur de masse qui fut récemment développé par Alexander Makarov [12, 13]. Les hautes performances en terme de résolution de masse pouvant être obtenues avec le LTQ-Orbitrap ainsi que ces avantages pratiques en comparaison au FT-ICR ont propulsés son utilisation à de vastes applications en protéomique.

L'analyseur de masse de type Orbitrap est une trappe électrostatique constituée d'une électrode creuse dans laquelle est placée une électrode centrale en forme de fuseau [13]. Les ions sont injectés tangentiellement à l'électrode centrale et piégés autour d'elle par l'application d'un voltage approprié avec les deux électrodes [14]. Le champ électrostatique contraint les ions à faire des mouvements complexes en forme spirale autour de l'électrode centrale. De la même façon que pour le FT-ICR, le courant induit de ces oscillations permet par l'intermédiaire d'une transformée de Fourier d'obtenir les valeurs m/z correspondantes [12]. L'Orbitrap est un instrument qui offre des précisions de masse inférieures à 2 ppm, une résolution ($M/\Delta M$) rivalisant avec celle du FT-ICR (jusqu'à $M/\Delta M$ = 100 000) ainsi qu'un domaine linéaire de l'ordre de 10^3 [12, 13]. L'analyseur de masse Orbitrap est par contre davantage utilisé comme détecteur étant donné que des pertes significatives de sensibilité sont observées dans le cadre d'analyses MS/MS. C'est d'ailleurs pour contrer cette lacune que l'hybride LTQ-Orbitrap fut conçu.

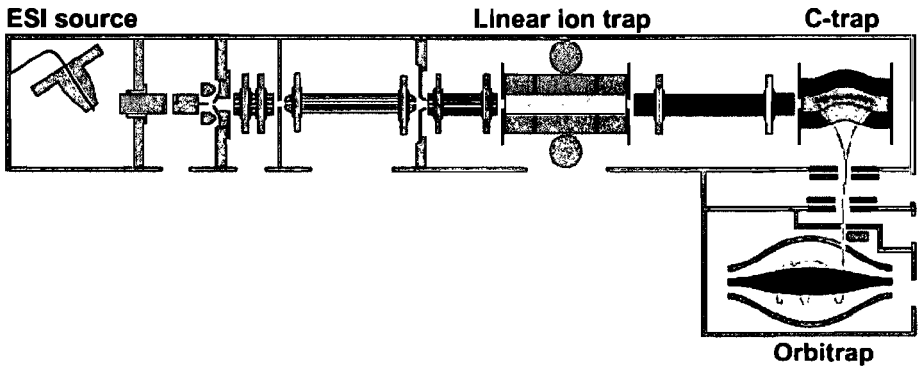

Figure 2.3: Représentation schématique du spectromètre de masse LTQ-Orbitrap. Reproduite de [13].

Le LTQ-Orbitrap est constitué de trois composantes principales : la trappe linéaire quadrupole (LTQ), la C-trap et l'Orbitrap (Figure 2.3). La LTQ est une trappe ionique bidimensionnelle dans laquelle les ions sont piégés au sein d'un champ quadrupôlaire. Le balayage de l'amplitude de la radiofréquence (RF) entraîne l'expulsion radiale des ions piégés vers le détecteur. Dans le cadre d'analyses MS/MS, seul l'ion peptidique étant stabilisé dans le champ électrique à un temps donné sera fragmenté au sein de la trappe [9]. La C-trap quant à elle est employée pour conserver et accumuler les ions préalablement à leur injection pulsée dans l'Orbitrap. Dans le cadre des expériences effectuées dans cette thèse, les deux analyseurs de masse furent utilisés en parallèle pour bénéficier de leurs avantages respectifs. La LTQ fut utilisée pour avoir une sensibilité et une rapidité d'acquisition maximale pour les spectres MS/MS. En parallèle, l'acquisition des spectres MS à haute résolution fut effectuée dans l'Orbitrap. Cette combinaison permet d'obtenir un grand nombre de spectres MS/MS (3-5 spectres/sec) tout en ayant une excellente exactitude de masse sur les ions peptidiques identifiés ($M/\Delta M$ = 60 000) [13]. La précision de masse est un critère très important pour limiter les fausses identifications surtout dans le cadre de l'étude du répertoire MIP où la correspondance avec la protéine source est souvent effectuée à partir d'une seule séquence peptidique.

2.3.3.1 Recherche dans les banques de données avec Mascot

Le moteur de recherche Mascot (http://www.matrixscience.com/) fut utilisé pour l'identification des spectres MS/MS expérimentaux dans les banques de données (section

1.3.2.4). Les paramètres de recherche sont fixés en fonction des caractéristiques de l'échantillon et du type d'analyse pour limiter le taux de fausses identifications. La figure 2.4 montre un exemple de la page de paramètres utilisée pour une recherche Mascot en mode MS/MS.

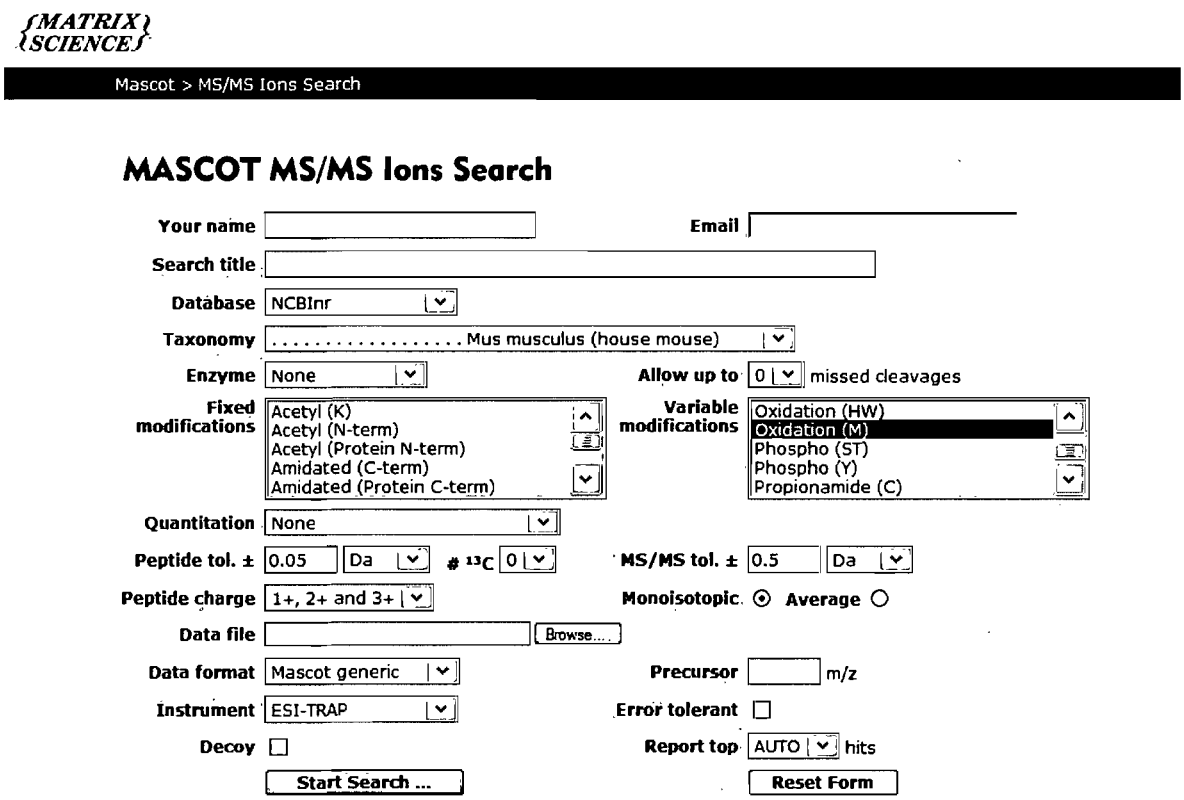

Figure 2.4: Page de paramètres pour une recherche Mascot en mode MS/MS. Seulement l'oxidation (M) en modification variable fut sélectionnée pour les analyses sur les extraits MHC I.

L'approche générale consiste à télécharger le fichier brut de données contenant les spectres MS/MS dans Mascot. On choisit ensuite la banque de données (e.g. International Protein Index, IPI; National Center for Biotechnology Information nonredundant, NCBInr) et les paramètres de recherche en fonction de la préparation (e.g. digestion enzymatique), du type d'échantillon (e.g. modifications) ainsi que par rapport au type d'instrument MS utilisé (Δm ions précurseurs et ions fragments). Mascot génère ensuite une liste d'identifications contenant l'information associée à chacun des spectres MS/MS par rapport à la séquence peptidique, la protéine d'origine, et les coordonnées de l'ion précurseur (e.g.

Rt, m/z, z). Les paramètres utilisés pour les recherches Mascot sur les extraits MHC I sont décrits dans la figure 2.4.

En protéomique, le clivage enzymatique trypsique permet d'avoir une plus grande spécificité sur la composition en C-terminal (lysine ou arginine) ce qui améliore la correspondance des spectres dans Mascot. De plus, la présence d'un acide aminé basique en C-terminal favorise la fragmentation le long de la chaîne peptidique permettant ainsi d'obtenir des spectres MS/MS de meilleure qualité [15]. Ce clivage enzymatique trypsique fut utilisé dans le cadre des travaux présentés dans le chapitre 3. Dans le cas de recherches effectuées sur des préparations MHC I, le taux de faux positifs (mauvaise correspondance séquence – spectre MS/MS) est souvent supérieur comparativement à des extraits de peptides trypsiques étant donné que les peptides MHC I ne dérivent pas d'un clivage enzymatique spécifique. Ces taux de faux positifs sont présentement évalués par la comparaison de recherches directes/inverses [16]. Une autre problématique est reliée au fait que les peptides MHC I sont de courtes chaînes peptidiques et que la fragmentation s'effectue de façon aléatoire en dépendance avec la composition en acides aminés basiques. Une validation manuelle des séquences peptidiques proposées par Mascot fut donc nécessaire pour l'identification des peptides MHC I afin de limiter le taux de faux positifs. Pour facilité cette tâche, un filtre informatique fut développé pour discriminer entre les peptides MHC I spécifiques et les autres peptides présentés dans les extraits (annexe AI). La sélection des peptides à l'aide du filtre fut effectuée en fonction de la spécificité MHC I (longueur des chaînes peptidiques et sites d'ancrages) et des motifs peptidiques attendus en lien avec les allèles étudiés ($H2K^b$, $H2D^b$ et Qa2, voir section 1.2.1).

2.4 Quantification relative du répertoire MIP

Plusieurs techniques de quantification relative par LC-MS furent développées en protéomique comme précédemment discuté dans la section 1.3.3. Dans notre laboratoire, la comparaison des profils d'expression est effectuée à l'aide de l'approche sans marquage (section 1.3.3 et section 1.4.2). Le développement de méthodes par nanoLC-MS robustes et reproductibles ainsi que d'outils bioinformatiques adaptés à l'analyse du protéome nous ont permis d'obtenir une plateforme protéomique plus efficace pour l'identification et la quantification de mélanges protéiques complexes. Cette méthodologie analytique suscite un

grand intérêt étant donné qu'elle est applicable à tous les modèles, qu'aucune modification n'est apportée à l'échantillon et que la quantification relative peut être effectuée sur un nombre infini de conditions expérimentales. Cette approche fut d'ailleurs utilisée pour la première fois à l'analyse différentielle du répertoire MIP dans le cadre des travaux présentés dans le chapitre 4 et 5. Les sections suivantes expliquent les grandes lignes du processus de traitement des données LC-MS à l'aide des différents logiciels développés au sein de notre laboratoire.

2.4.1 Identification des peptides

La première étape essentielle au succès de l'approche sans marquage est d'avoir un logiciel efficace pouvant faire l'identification des peptides. Pour ce faire, le logiciel Mass Sense fut développé pour identifier chacun des peptides présentés à travers les fichiers de données brutes [17]. Ce logiciel permet d'attribuer avec un taux de confiance élevé les coordonnées de chacun des peptides en terme de valeur d'abondance, de m/z, de charge (z) et de temps de rétention (Rt) (Figure 2.5). Ces coordonnées sont définies par le logiciel à l'aide d'une modélisation peptidique des patrons isotopiques (Figure 2.5). Cette signature isotopique rend possible l'identification par Mass Sense des coordonnées m/z, Rt et z pour l'intensité maximale de la masse monoisotopique ^{12}C de chacun des ions peptidiques.

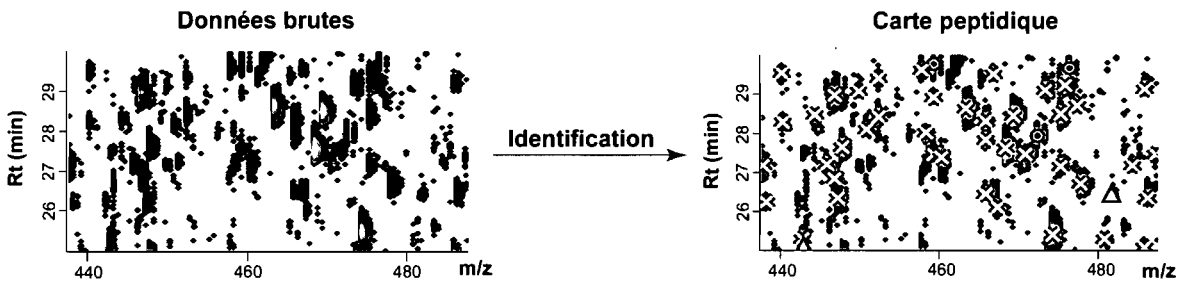

Figure 2.5: Profils générés par Mass Sense pour les données brutes avant et après l'identification des peptides. L'abondance des peptides est représentée à l'aide d'une échelle de couleurs de noir (faible abondance) à jaune (abondance élevée). Les cercles, les croix et les triangles correspondent à des espèces simplement, doublement ou triplement chargés, respectivement.

La confiance dans l'identification des peptides est établie à l'aide d'un pointage qui est attribué à chacun des ions peptidiques permettant d'indiquer le taux de concordance entre le patron isotopique observé et celui attendu. Le logiciel permet de générer des listes

en format excel comportant les coordonnées de chacun des peptides (m/z, Rt, z, abondance, pointage). Des analyses de segmentation doivent ensuite être effectuées à l'aide du logiciel Peptide Clusterer pour faire la quantification relative entre les conditions expérimentales (voir section suivante 2.4.2). Une validation manuelle du logiciel a permis d'établir que l'identification fait par Mass Sense contenait un taux de faux négatifs (peptides présentés mais non identifiés) de 7,4% et un taux de faux positifs (fausses identifications) de 4,6% [17].

2.4.2 Analyses de segmentation

Les analyses de segmentation permettent la comparaison de l'abondance des ions peptidiques à travers différentes conditions expérimentales. Les cartes peptidiques générées par Mass Sense sont comparées à l'aide du logiciel Peptide Clusterer pour regrouper les peptides à travers les réplicats et les différentes conditions biologiques selon des tolérances fixées au niveau des m/z, des temps de rétention et des fractions (2D-LC) (Figure 2.6). Ces tolérances sont fixées en fonction de la résolution de masse de l'instrument utilisé ainsi que par rapport à la reproductibilité des analyses (e.g. Orbitrap : m/z ± 0.02, Rt ± 1 min).

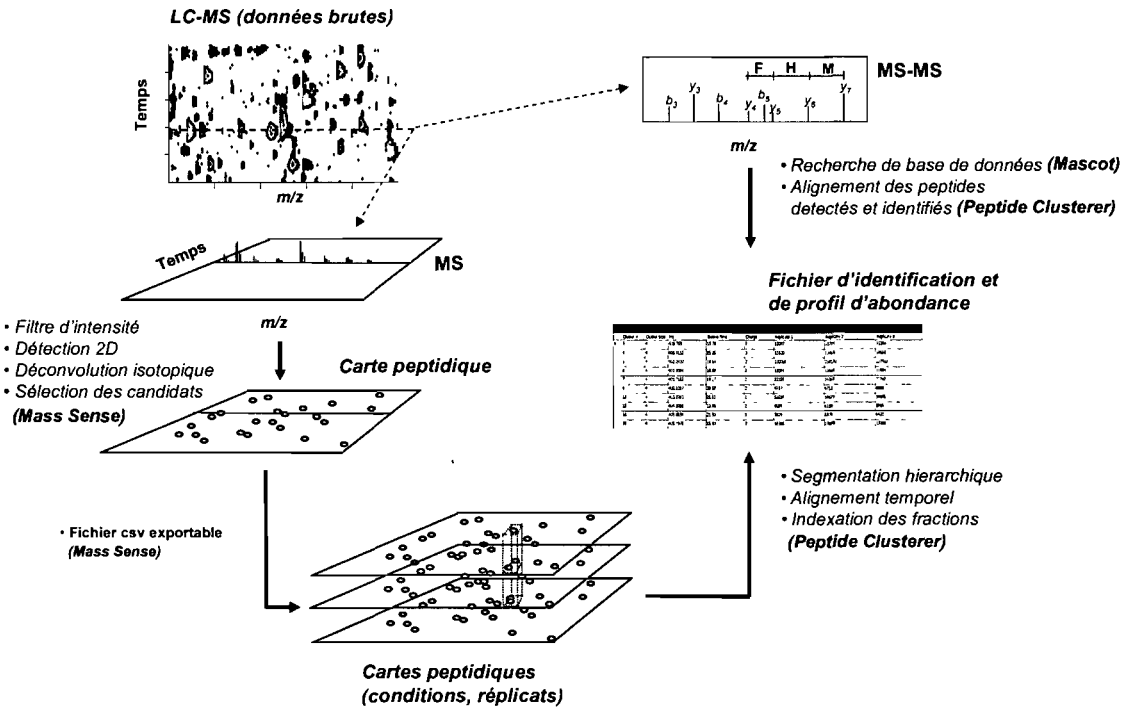

Figure 2.6: Processus de traitement des données à l'aide des logiciels Mass Sense, Peptide Clusterer et Mascot. Modifiée de [17].

Ce processus permet de générer une liste de groupe de peptides comportant les intensités pour chacune des conditions expérimentales facilitant ainsi l'analyse différentielle de mélanges complexes. Chaque ligne de cette liste est associée à un ion peptidique unique dont les coordonnées (Rt, m/z, z, intensité) sont répertoriées pour chacune des conditions expérimentales. Cette liste de groupe de peptides peut ensuite être alignée avec la liste d'identification provenant de la recherche Mascot. Ces étapes combinées permettent donc d'identifier et de quantifier les changements d'abondance à travers les conditions expérimentales étudiées (e.g. normal vs. cancer). La figure 2.6 résume le processus de traitement de données à l'aide des différents logiciels.

2.5 Références

1. Purcell, A.W., *Isolation and characterization of naturally processed MHC-bound peptides from the surface of antigen-presenting cells.* Methods Mol Biol, 2004. **251**: p. 291-306.

2. Fortier, M.H., Caron, E., Hardy, M.P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P., *The MHC class I peptide repertoire is molded by the transcriptome.* J Exp Med, 2008. **205**(3): p. 595-610.

3. Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., and van de Goor, T., *Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip.* Anal Chem, 2005. **77**(2): p. 527-33.

4. Belder, D. and Ludwig, M., *Surface modification in microchip electrophoresis.* Electrophoresis, 2003. **24**(21): p. 3595-606.

5. Bandilla, D. and Skinner, C.D., *Capillary electrochromatography of peptides and proteins.* J Chromatogr A, 2004. **1044**(1-2): p. 113-29.

6. Shen, Y., Moore, R.J., Zhao, R., Blonder, J., Auberry, D.L., Masselon, C., Pasa-Tolic, L., Hixson, K.K., Auberry, K.J., and Smith, R.D., *High-efficiency on-line solid-phase extraction coupling to 15-150-microm-i.d. column liquid chromatography for proteomic analysis.* Anal Chem, 2003. **75**(14): p. 3596-3605.

7. Ghitun, M., Bonneil, E., Fortier, M.H., Yin, H., Killeen, K., and Thibault, P., *Integrated microfluidic devices with enhanced separation performance: application to phosphoproteome analyses of differentiated cell model systems.* J Sep Sci, 2006. **29**(11): p. 1539-49.

8. Fortier, M.H., Bonneil, E., Goodley, P., and Thibault, P., *Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs.* Anal Chem, 2005. **77**(6): p. 1631-40.

9. Watson, J.T. and Sparkman, D.O., *Introduction to Mass Spectrometry: Instrumentation, Applications ans Strategies for Data Interpretation.* 4 ed. 2007, John Wiley & Sons, Chischester, England: p. 53-171.

10. Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., 3rd, *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.

11. Washburn, M.P., Wolters, D., and Yates, J.R., 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.

12. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., and Horning, S., *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer.* Anal Chem, 2006. **78**(7): p. 2113-20.

13. Scigelova, M. and Makarov, A., *Orbitrap mass analyzer--overview and applications in proteomics.* Proteomics, 2006. **6 (Suppl 2)**: p. 16-21.

14. Makarov, A., Denisov, E., Lange, O., and Horning, S., *Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer.* J Am Soc Mass Spectrom, 2006. **17**(7): p. 977-82.

15. Kinter, M. and Sherman, N.E., *Protein Sequencing and Identification Using Tandem Mass Spectrometry.* 2000, John Wiley & Sons, New York, NY: p. 64-80.

16. Elias, J.E. and Gygi, S.P., *Target-decoy search strategy for increased confidence in large-scale protein identifications by MS.* Nat Methods, 2007. **4**(3): p. 207-214.

17. Bonneil, E., Jaitly, G., Jaitly, N., Pomiès, C., and Thibault, P., *Comprehensive expression profiling and trace-level identification of unlabeled peptide ions in 2D-LC-MS proteomics experiments using integrated detection and clustering software*, Proceedings of 55th ASMS conference on Mass Spectrometry & Allied Topics, June 3-7, 2007, Indianapolis, IN

3. Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs

Marie-Hélène Fortier, Eric Bonneil, Paul Goodley, and Pierre Thibault

Analytical Chemistry, 2005. 77(6): p. 1631-1640

3.1 Abstract

The present investigation describes the analytical performances of a microfluidic device comprising an enrichment column, a reversed-phase separation channel and a nanoelectrospray emitter embedded altogether in polyimide layers. This configuration minimizes transfer lines and connections and reduces postcolumn peak broadening and dead volumes. This compact and versatile modular nanoLC-chip system was interfaced to both ion trap and time-of-flight mass spectrometers and its analytical potential was evaluated in the context of proteomics applications. The figures of merit of this system in terms of peak capacity, reproducibility, sensitivity and linear dynamic range of peptide detection were determined using tryptic digests of complex protein extracts including albumin and immunoglobulin-depleted rat plasma samples. The analysis of peak profiles for more than 600 peptide ions reproducibly detected across replicate nanoLC-chip-MS runs (n=10) indicated that this system provided good reproducibility of retention time and peak intensity with RSD values of less than 0.5 and 9.1 %, respectively. Variation in peptide abundance as low as two-fold changes were identified for spiked tryptic digests present at levels of 2-5 fmol in plasma samples. Sensitivity measurements were performed on dilution series of protein digests spiked into rat plasma samples and provided a detection limit of 1-5 fmol. The modular concept of the microfluidic system also facilitated the integration of two-dimensional chromatography (strong cation exchange/C_{18}) thereby increasing the sample loading and selectivity of the nanoLC-chip-MS system. The application of this integrated device was evaluated for complex rat plasma samples to compare the number of protein identifications obtained using one- and two-dimensional nanoLC-chip-MS/MS.

3.2 Introduction

The purpose of proteomics goes beyond the identification of proteins in an organelle or in biological fluids. The need to understand the biological mechanisms involved in cancer, infectious, and inflammatory diseases at the clinical levels implies not only comprehensive protein identification but also expression profiling of proteins across healthy and disease patient samples. In this context, the challenge of proteomics lies in the complexity of protein mixtures, the number of samples, and the reproducibility of analysis. Most popular mass spectrometry-based proteomic approaches [1, 2] rely on two major strategies: (i) the traditional protein gel-based separation followed by capillary liquid chromatography-mass spectrometry (LC-MS) analyses of tryptic digests of 1D bands or 2D spots or (ii) the shotgun proteomic approach (2D-LC-MS) combining strong cation exchange (SCX) separation of tryptic digest followed by capillary LC-MS analyses of the salt fractions. Both techniques have shown to be very effective for protein identification [3-6] and expression profiling [7-9].

Capillary LC-MS systems used in those two strategies consist of an enrichment precolumn, an actuated rotor valve, and a nano-scale C_{18} reversed-phase analytical column terminated by a transfer line to the nanoelectrospray emitter. In systems operating at sub-microliter/min flow rate, transfer lines and dead volume must be minimized significantly in order to preserve separation efficiencies. Physical constraints with instrumentation and the requirements for routine operation and avoidance of capillary leakages and clogging often force the user to trade separation performance for ease of operation and maintenance. In addition, the serial nature of current capillary LC-MS systems limits the sample throughput and potential application for multiplex analyses although recent reports have described the successful implementation of a multiplex LC-MS system for multi-dimensional liquid chromatography separations [10]. In this regard, a fully integrated system comprising desalting and separation columns terminated by a short nanoelectrospray emitter would offer significant advantages over present nanoLC-MS configurations. Furthermore, miniaturization of the separation channel increase sensitivity due to greater concentrations of analyte bands eluting from smaller inner diameter columns [11, 12]. Higher demands for increased throughput, sensitivity and robustness from the pharmaceutical industry has led to

the integration and miniaturization of sample separation systems as well as coupling of microdevices with electrospray ionization mass spectrometry [13, 14].

For practical reasons, separation of peptides on microfluidics devices have been historically performed using systems based on capillary electrophoresis (CE) [15] and more recently using capillary electrochromatography [16]. A number of reports have presented the coupling of CE-based microfluidics devices to nanoelectrospray mass spectrometry [17-21]. On-line stacking or adsorption preconcentration prior to electrophoretic separation were also developed to enhance sensitivity and to decrease concentration detection limits to low-nanomolar levels [22, 23]. Significant advances in automation are still needed to achieve the required throughput and ruggedness sought for the rapid and reproducible proteomics analyses. To this end, efforts are also devoted to microfabricated chip designs enabling sample multiplexing [21, 24].

On the other hand , microchip LC-MS separations have been first implemented using either microchip wall derivatization with a stationary phase [25, 26] or UV light-induced polymerization of monolithic materials into the microchannel [27, 28]. Although those two approaches do not suffer from clogging and backpressure, the reduced adsorption area often results in lower loading capacity than traditional bead stationary phase. In the present work, we investigated the analytical performances of a newly designed nanoLC-chip-MS device for proteomic analyses [29]. This compact polyimide microfluidic device comprises laser-ablated channels embedding C_{18}-packed desalting and analytical separation channels together with a nanoelectrospray tip. The chip is mounted on a high-pressure manifold that consists of a face-sealed rotary valve that allows switching between sample desalting and separation while minimizing time delays and dead volumes. Application of this LC-chip-MS device using one- and two-dimensional separations is demonstrated for expression profiling and protein identification of rat plasma samples using both ion trap and time-of-flight (TOF) mass spectrometers.

3.3 Experimental section

3.3.1 Chemicals and materials

Fused-silica capillaries were purchased from Polymicro Technologies (Phoenix, AZ). Teflon and PEEK tubing were obtained from Supelco (Bellefonte, PA). HPLC grade water and acetonitrile (ACN) were purchased from Fisher Scientific (Whitby, ON, Canada). Formic acid (FA) and ammonium acetate were obtained from EM Science (Mississauga, ON, Canada). Protein standard digests used in the analyses were purchased from Michrom Bioresources (Auburn, CA). Rat plasma sample was obtained from Bioreclamation (Hempstead, NY). The Jupiter C_{18} 5 μm particle material used for packing the microfluidic devices was purchased from Phenomenex (Torrance, CA). The nanoLC-Chip and interface system were provided by Agilent Technologies (Santa Clara, CA). Strong cation exchange material was purchased from PolyLC (Columbia, MD).

3.3.2 Sample preparation

A solution of 8 protein standards (40 fmol/μl each) was prepared from individual digests of bovine serum albumin, rabbit phosphorylase *b*, yeast alcohol dehydrogenase, bovine deoxy-ribonuclease, horseradish Peroxidase C1A, bovine glyceraldehyde 3P dehydrogenase, *E. coli* beta-galactosidase and bovine carboxypeptidase A (Michrom Bioresources, Auburn, CA) diluted in a H_2O/5% ACN/0.2% FA solution. Complex protein extracts from rat plasma (Bioreclamation, Hempstead, NY) were depleted with IgG (Protein G, Amersham) and albumin (MP Biomedicals) columns to remove these high-abundance proteins. The rat plasma sample was digested in 0.1 M urea, 50 mM ammonium bicarbonate with Promega Trypsin (Fisher Scientific, Whitby, ON, Canada) overnight and dissolved in a H_2O/5% ACN/0.2% FA solution after being evaporated to dryness.

3.3.3 Microchip device fabrication

The microchip device fabrication is described elsewhere [29]. A polyimide film (6.5 × 2.5 cm) is laser-ablated to form microfluidic channels, ports, chambers and columns. The

ablation residues are removed and the film is laminated with heat and pressure application. The chip is then trimmed using laser ablation to its final shape and to form the electrospray tip. The final 8 μm i.d. tip shape is conical with a circular end with an emitter of 100 μm o.d. and 2 mm in length (Figure 3.1). A gold layer is electrodeposited near the electrospray tip for contacts to the fluid flow channel. The layer surface bearing the electrical contacts is directly laminated such that the metal is embedded inside the device. The enrichment precolumn and the LC column were packed with 5-μm Jupiter C_{18} material. The internal volume of the enrichment precolumn is 40 nL whereas the LC column has a 75 by 50 μm cross section and a length of 4.5 cm. The analytical column is connected to the tip via a 15 μm i.d. × 0.4 mm channel inside the chip.

3.3.4 Chip valve manifold

The chip valve manifold is described in more detail elsewhere [29]. Briefly, the chip is positioned between the stator and the rotor of a two-position Rheodyne HPLC valve. A clamping arm mechanism ensures that the stator ports are aligned to the ports on top of the chip, and the rotor channels to ports are aligned via registration pins. When the clamping arm is closed, the valve stator, the chip, and the valve rotor are sealed together. The arm is rotated to position the chip orthogonally to the sampling cone of the mass spectrometer.

3.3.5 NanoLC-chip-MS and nanoLC-chip-MS/MS analyses

An HP1100 capLC pump (Agilent Technologies) was used to deliver solvent H_2O/3% ACN/0.2% FA during sample loading and an Agilent nanoLC 1100 Series was used for gradient delivery. During sample loading the flow rate was set to 4 μL/min for 6 min. LC gradient (mobile phase A: H_2O, 0.2% FA; mobile phase B: ACN, 0.2% FA) was delivered by an Agilent 1100 nanoPump with flow rate set at 300 nL/min. Peptides were eluted from the reversed-phase column into the mass spectrometer using a linear gradient elution of 10% to 60% B over 56 min. The nanoelectrospray voltage was set to 2400 V. All nanoLC-chip-MS experiments were performed on a TOF and a ion Trap XCT (Agilent Technologies) mass spectrometers.

3.3.6 Two-dimensional LC separations with on-line LC-chip MS/MS analyses

The SCX column (0.5 mm i.d. × 6.3 cm) was packed with PolyLC stationary phase (5 μm diameter, 300 Å pore size). The SCX column was connected directly to the switching valve, and was on-line with the chip C_{18} precolumn during sample loading, and toggled off-line during reversed-phase peptide separation on the analytical column. Peptides were sequentially eluted from the SCX column onto the C_{18} precolumn with 8 salt fractions of ammonium acetate, pH 3.5. Each 10 μL salt fraction was loaded from the autosampler at 4 μL/min for 15 min and delivered to the SCX cartridge using an auxiliary pump. Peptides were eluted from the SCX column using sequential fractions of 0, 55, 65, 80, 100, 150, 400 and 1000 mM ammonium acetate, pH 3.5. Peptides were eluted from the reverse phase column into the mass spectrometer using a gradient from 10% to 60% B over 56 min. The nanoelectrospray voltage was set to 2400 V. Tandem mass spectra were acquired with the Agilent MSD Trap XCT mass spectrometer using He as a collision gas. Multiply charged ions with an intensity above a threshold of 40,000 counts were selected for MS/MS sequencing. The MS/MS fragmentation amplifier voltage was set to 1.3V.

3.3.7 Peptide detection and clustering

A script was developed to convert the raw data files from the ion trap or TOF instruments into text files representing all m/z, intensity and time values above a user-defined intensity threshold, typically 100 counts for TOF and 700 000 counts for ion trap data. These text files were then processed using in-house softwares enabling data reduction, peptide detection and alignment of peptide ion maps (m/z, retention time, intensity) across all samples [30]. Segmentation analyses were performed using hierarchical clustering with criteria based on m/z and time tolerance (typically ± 0.1 m/z and ± 1 min) to generate lists of non-redundant peptide clusters for all replicates or samples to be compared.

3.3.8 Protein identification

Database searches were performed against a non-redundant NCBI database using Mascot (Matrix Science, London, UK) selecting human and rodent species. Parent ion and fragment ion mass tolerances were both set at +/- 0.4 Da.

3.4 Results and discussion

This work focuses on the evaluation of the analytical performances of a nanoLC-chip system for the analysis of highly complex peptide mixtures. A photograph of this microfluidic device interfaced to an Agilent ion trap XCT is shown on Figure 3.1. The inset shows a close up of the device integrating together a 40 nL volume precolumn, a 4.5 cm length analytical column with a 75 x 50 µm cross section channel and a nanoelectrospray tip. The chip is placed on a support plate housing a valve rotor on one side and aligned inlet ports on the other side. The valve stator is mounted in a clamping mechanism enabling proper alignment and sealing of the chip within the manifold. The chip, support plate and, rotor valve are mounted on an adjustable two-way translation stage via a rotating bracket allowing easy positioning of the manifold in front of the mass spectrometer sampling cone. When in position, the nanoelectrospray tip is perpendicular and approximately 2.5 mm axially and 1 mm off-axis from to the inlet cone. Precise positioning of the spray emitter is facilitated through a camera and lighting system.

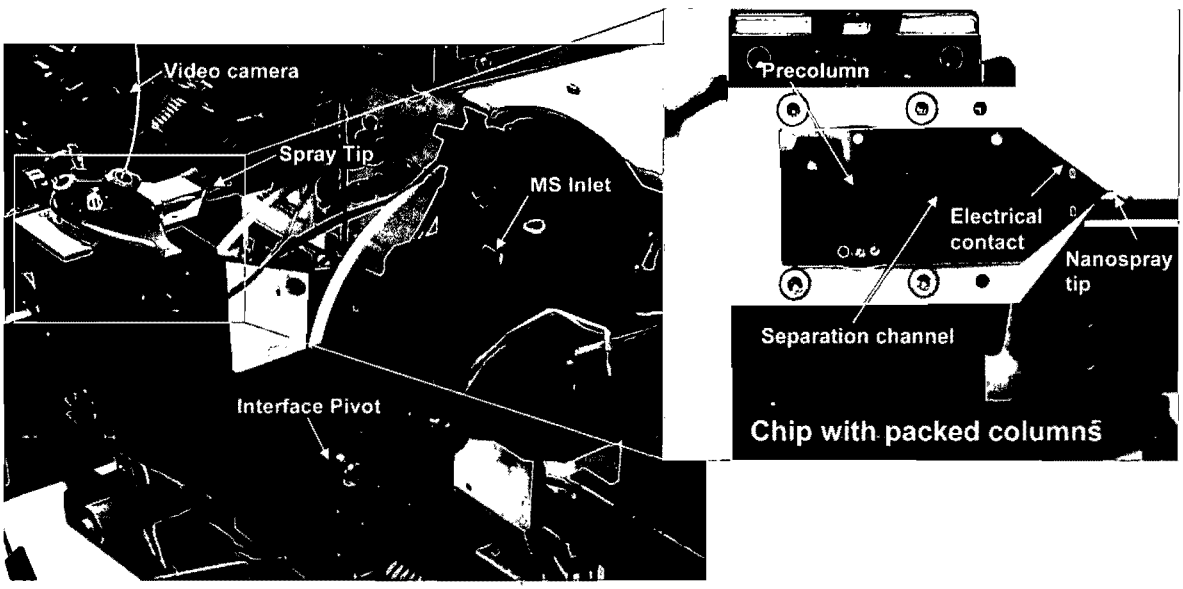

Figure 3.1: NanoLC-chip-MS system interfaced to an Agilent ion trap XCT mass spectrometer. The chip rests on an articulated manifold comprising a valve rotor with a clamping mechanism ensuring proper port alignment and sealing. A two-way translation stage provides easy positioning of the device in front of the sampling orifice when the manifold is inserted in the MS inlet. The inset shows a close-up view of the chip device with the precolumn (40 nL), separation channel (4.5 cm length x 75 µm x 50 µm cross section channel), and nanospray tip (8 µm i.d.).

3.4.1 Analytical performances of the LC-chip-MS system

The application of this microchip LC device coupled to the time-of-flight mass spectrometer (nanoLC-chip-TOF) for proteomics analyses was first tested using a tryptic digest of an 8-protein mixture representing a diverse population of peptides with different molecular masses and hydrophobicities. The base peak chromatogram corresponding to the injection of 80 ng of the combined digest (200 fmol of each protein digest) is presented in Figure 3.2a along with the contour profile of m/z vs time vs intensity in Figure 3.2b.

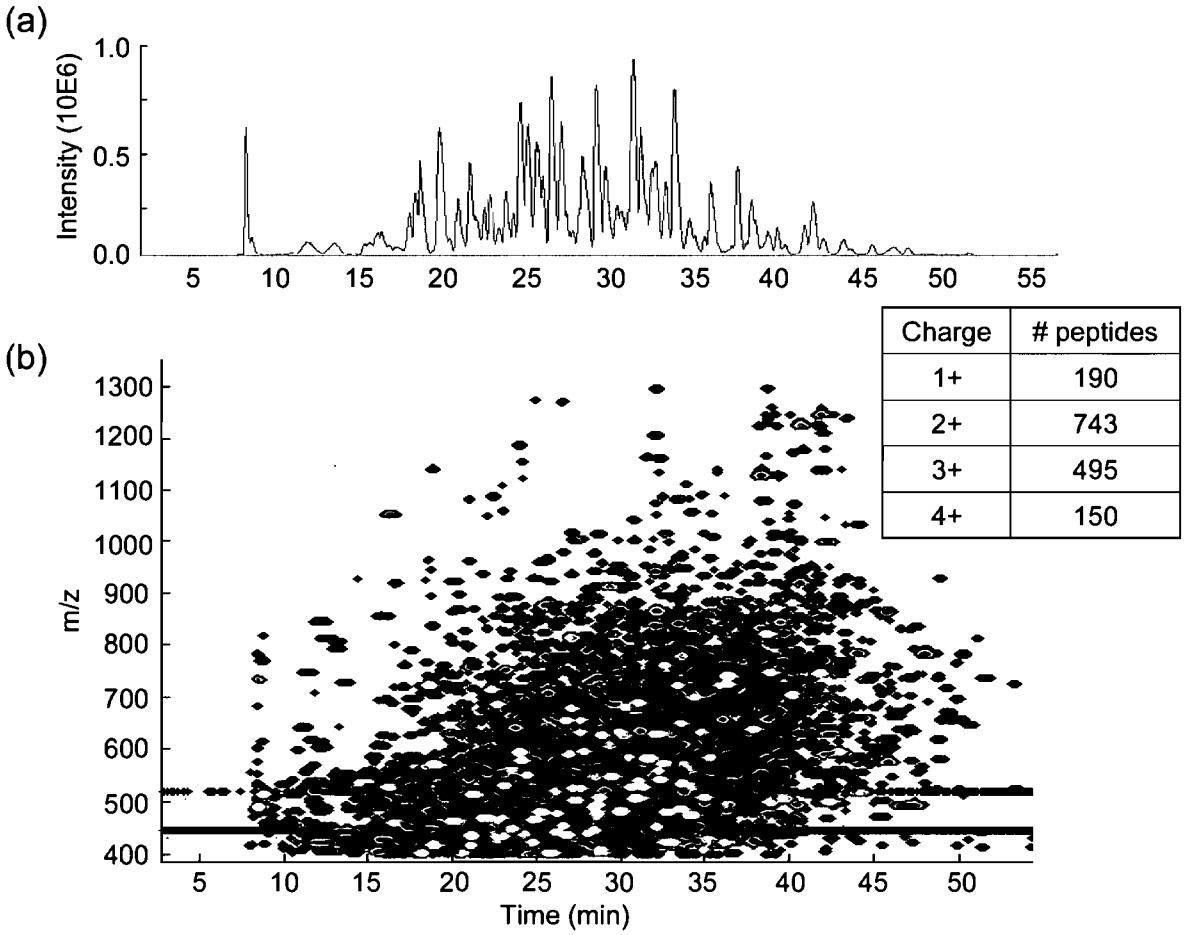

Figure 3.2: NanoLC-chip-LC-MS analysis of a 80 ng injection of an 8-protein tryptic digest. (a) Base peak chromatogram. (b) Contour profile of m/z *vs* retention time *vs* intensity. Peptide ions are displayed on a logarithmic scale of increasing intensity ranging from black to bright yellow. The number of detected peptide ions and their corresponding charge state are shown in the table.

The logarithmic intensity scale shows an increase in abundance for ions transitioning between black and bright yellow. Each dot shown on this contour profile

represents an individual peptide ion defined by specific m/z and time coordinates. Peptide detection was facilitated using an in-house proprietary software program that analyze the peak profiles of all eluting isotopic components to return ^{12}C monoisotopic peptide ions based upon their corresponding *m/z* value, time, charge state and intensity [30]. The charge distribution of the 1578 detected peptides is shown in the table of Figure 3.2a. Comparison of peptide maps across replicates or sample sets is thus facilitated by digitizing the nanoLC-chip-MS analysis to produce lists of detected peptide ions.

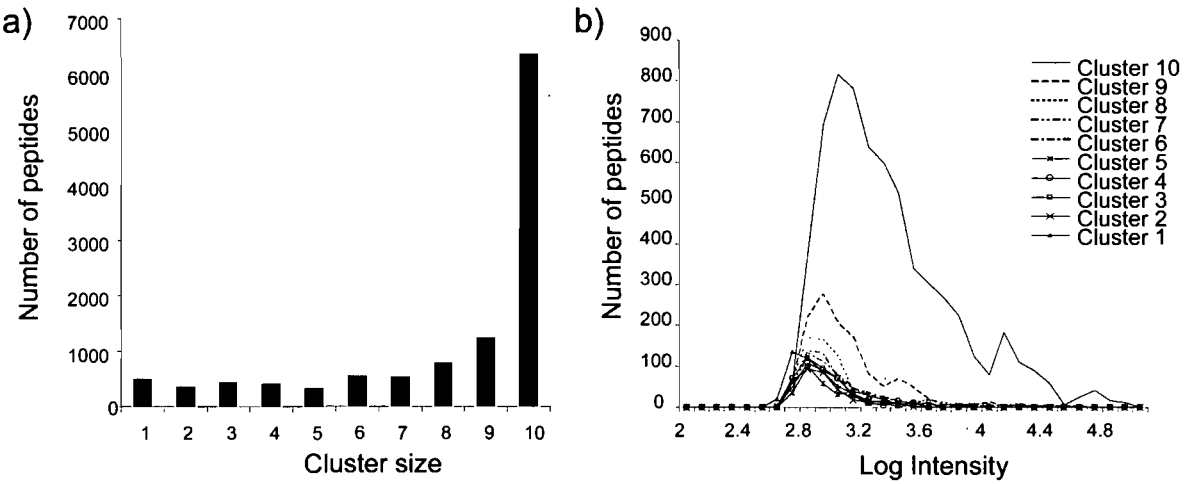

Figure 3.3: Reproducibility of peptide detection across 10 replicate nanoLC-chip-MS analyses of an 8-protein tryptic digest. (a) Number of peptide ions found across replicates or (a) cluster size and (b) log of ion intensity.

To determine the reproducibility of elution profiles in terms of time, m/z measurements and intensity, we compared the digitized peptide maps of 10 replicate analyses of the 8-protein digest mixture. The identification of individual peptide cluster across replicates (defined as unique ions reproducibly found within +/- 0.1 m/z and +/- 1.0 min) was achieved using segmentation analyses with hierarchical clustering. These unique peptide clusters were then compared to determine their frequency of observation across all 10 replicates. Figure 3.3a represents the distribution of cluster size defined as the number of times a given peptide cluster was detected across the 10 replicates. The number of peptides was normalized to include all individual peptides found in a specific cluster size. From a total number of 11 456 detected peptide ions, 55 % of this peptide population was observed in all 10 replicates whereas 83 % was detected in at least half of the entire peptide population (cluster size $\geq$ 6). It is noteworthy that this comparison was performed across all

peptide ions irrespective of their individual intensities and does not account for the higher variability in peptide detection typically observed for lower intensity ions. Indeed, examination of the peptide ion population represented by cluster size 1 indicates a higher proportion of lower abundance ions when compared to higher cluster sizes (Figure 3.3b). For example, the median value of ion intensity for cluster size 1 was 594 counts (log intensity, 2.77) compare to 707 counts (log intensity, 2.85), and 1480 (log intensity, 3.17) for cluster sizes 6 and 10, respectively. Obviously, low-abundance peptides are more difficult to detect reproducibly across numerous replicates due to factors such as retention competition on the precolumn and analytical column, background variation, and ionization suppression.

To determine the reproducibility of observed retention times, m/z values, and intensities we compared the distribution of relative standard deviations (RSD) across peptide clusters. Figure 3.4a gives the relative standard deviation of retention times for 626 peptide clusters found in all 10 replicates. As indicated RSD values varied typically between 0.08 and 1.05 % with a median value of 0.27 %. Furthermore, 95% of peptide ions reproducibly detected across all replicates showed a variation in retention time below 0.15 min corresponding to an RSD of less than 0.5 %. Similarly, a comparison of m/z measurements indicated that individual peptide clusters were represented by RSD values ranging from 0.001 to 0.003 %, and 95 % of all reproducibly detected ions were within +/- 25 ppm. It is noteworthy that these m/z measurements were performed with external calibration and improvement in mass accuracy to less than 5 ppm could be achieved by co-infusion of an internal standard.

The reproducibility in ion intensity observed across peptide clusters of size 10 is shown in Figure 3.4b which compares the log-log plot of individual vs average intensities. Most of the ions are tightly distributed along a 45° angled line for all replicates indicating a narrow dispersion across 2.5 orders of magnitude in intensity. As observed from Figure 3.4b, intensity variations appeared more significant for lower abundance ions with an average RSD value of 9.12 % for ions of less than 2000 counts compared to an average RSD value of 6.19% for ions above this intensity level. An expression plot of fold change variation in intensity for all peptide clusters of size 10 (inset Figure 3.4b) highlights that 95% of peptide clusters showed a variation of less than ± 20% (± 1.2 in fold change) across

replicate injections, attesting of the good reproducibility of the present system. Reproducibility of retention time and peak intensity of the nanoLC-chip-TOF system is comparable to that observed previously on a nanoscale capillary LC system coupled to a Q-Star mass spectrometer [12].

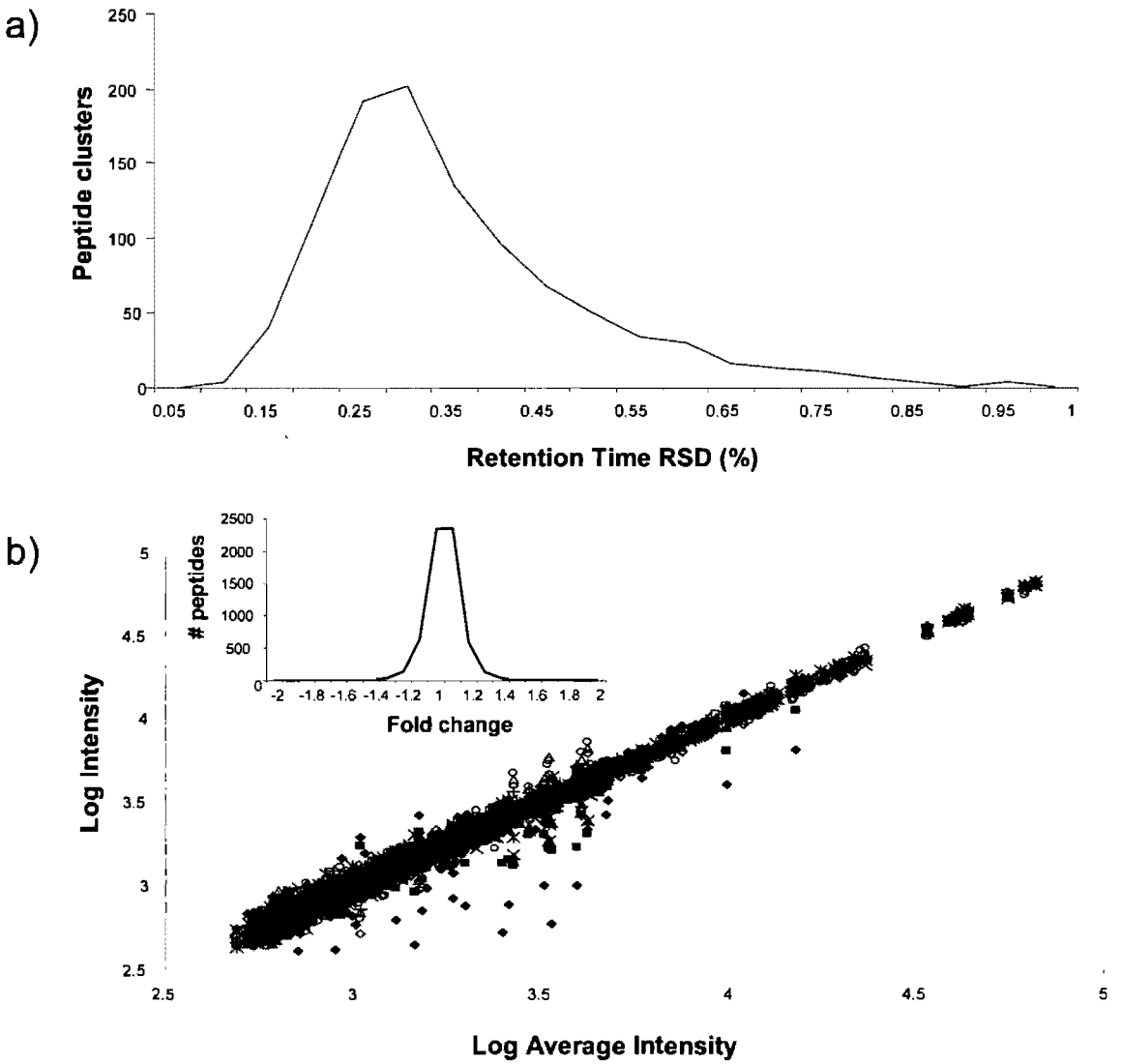

Figure 3.4: Reproducibility of retention time and ion intensity for peptides reproducibly found in all 10 replicate injections of an 8-protein tryptic digest. (a) Distribution of RSD values for all of the 626 individual peptide clusters. (b) Log-log plot of individual ion intensity *vs* the average intensity of the peptide cluster. Inset shows the frequency of peptide detection across the fold change in ion intensity.

The evaluation of peak capacity and peak asymmetry was performed on peptide clusters reproducibly observed across all 10 replicates. Peak width measurements were

performed at 10 and 50% height for detected ions. An example of a peak profile for the doubly charge peptide ion at m/z 501.72 is shown in Figure 3.5a. The distribution of peptide clusters as a function of peak widths at 50% and 10% is presented in Figure 3.5b and data obtained are summarized in Table 3.1. The wider distribution observed for peak width at 10 % is due to larger variability in intensity values for measurements taken at levels approaching background noise. The mean values for peak widths measured at 50 and 10 % height were 0.34 and 0.76 min, respectively (Table 3.1). In comparison, packed capillary columns of 12 cm x 75 μm i.d. gave average peak widths of 0.27 and 0.65 min for measurements taken at 50 and 10% height (data not shown).

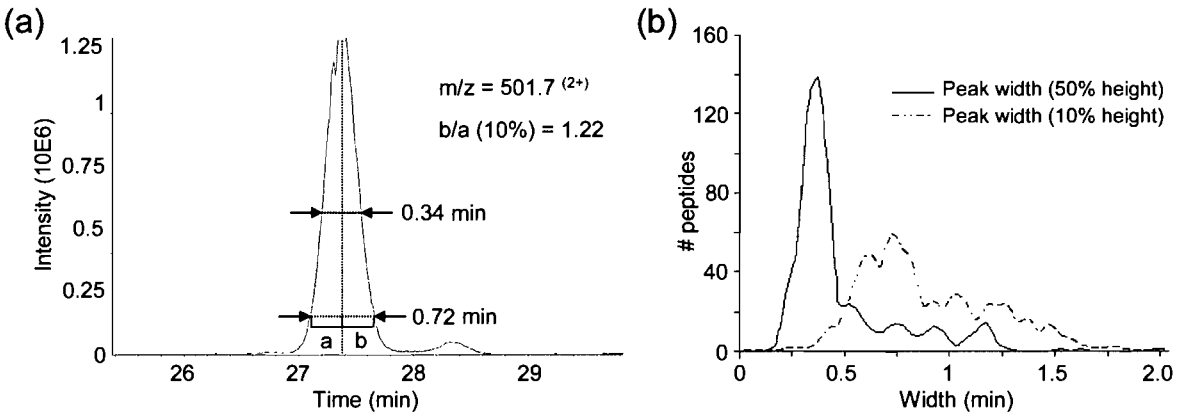

Figure 3.5: Peak asymmetry and peak width measurements on the nanoLC-chip-MS system. (a) Extracted ion chromatogram on a doubly protonated peptide at m/z 501.7 eluting in the most dense region of the contour profile shown in Figure 3.2b. (b) Distribution of peptide ions according to peak widths at 50 and 10 % height.

Peak asymmetry was also measured for a set of 240 reproducibly observed peptides with intensity above 10 000 counts to minimize variability attributed to poor peak definition at levels approaching background noise. The ratio of half width along the center peak line, b/a (figure 3.5b) gave median peak asymmetry values of 1.13 and 1.22 for measurements taken at 50 and 10% height, respectively (Table 3.1). In contrast, peak asymmetry values of 1.49 and 1.93 were obtained for packed capillary columns of 12 cm x 75 μm i.d. for measurements taken at 50 and 10% height (data not shown). The low asymmetry values obtained for the microfluidic device were attributed to low dead volumes and reduced postcolumn peak-broadening contributions. Correspondingly, the peak capacity values, n_c, were calculated at 50% height and 10% height from the peak width

measurements (Table 3.1) and provide an estimate of the number of components that can be separated to a specified resolution within a given separation space [31]. An average n_c value was calculated from the ratio of the elution time window and the peak width as described previously [12, 32]. The modest average n_c value of 141 (50 % height) compared to other reports using 70 cm x 50 μm id column [12] reflects the use of an on-line trapping column, the narrower elution window, and the difference in particle size used. Indeed, increased peak capacity could be obtained by using shallower gradient elution for a given column configuration. However, more appreciable gains in separation performance are expected by reducing the precolumn volume, increasing the column length and using smaller particle size.

Table 3.1: Average peak width and asymmetry measurements obtained from nanoLC-chip-MS analyses of tryptic peptides from an 8-protein digest.

% Height	Peak width (min)	Peak capacity (n_c)	Asymmetry (b/a)*
50	0.34	141	1.13
10	0.76	63	1.22

* Statistics across 240 peptides (n =5) with intensity higher than 10 000 counts

Another important consideration is the ruggedness and ease of operation of the present system. Despite the reduced dimensions of the precolumn and analytical columns compared to more conventional capillary LC systems, the integrated format of this device enables a significant reduction in dead volumes while simultaneously minimizing leakage through tubing connections. These advantages were translated by reduced peak asymmetry compared to capillary columns of comparable size. Although the robustness of this microfluidic system was not necessarily the subject of the present investigation, the nanoLC-chip-MS system was operated in an uninterrupted manner for more than 100 injections of complex proteins digests on a single chip without any noticeable degradation of performance. Moreover, the articulated manifold facilitates chip replacement and system reconditioning within less than five minutes.

3.4.2 Detection and identification of spiked proteins in a blood biomarker sample

Replicate analyses described in the previous section highlighted the good reproducibility of the present system where RSD values of less than 0.5, 0.003, and 9.1 % were obtained on retention time, m/z value, and intensity measurements, respectively. These reproducibility measurements provided comfortable operating margins for peak detection across comparable samples. To evaluate the analytical potentials (e.g. reproducibility, sensitivity, limit of detection) of the present microfluidic device in the context of expression profiling measurements, we performed a control experiment on complex rat plasma samples. In this case, each sample injected on the chip corresponded to different levels ranging from 0 to 250 fmol of the 8-protein digest spiked into a 100 ng digest of plasma sample. Spike series were injected in triplicate interleaving three blank samples before the next series of injections. Peptide maps and clustering analyses were performed as described in the previous section.

The average number of reproducibly detected peptide ions in all three nonspiked replicate samples was 992 ions. The distribution of peptide ions unique to the plasma sample and reproducibly found in the spike series was relatively uniform and ranged from 931 to 976 ions. A comparison of peptide ions reproducibly detected and specific to the spiked samples progressively increased from 4 to 604 ions for spike levels of 1.2 to 120 fmol. As expected, changes in peptide intensity observed at the low-femtomole level were observed for tryptic peptides showing a higher signal response. To visualize the intensity changes of tryptic peptide ions from spiked proteins we compared the log-log plots of the 24-, 48- and 120-fmol spike with that of the 12 fmol (Figure 3.6a). At the lower spike level, we identified a total of 61 unique peptide ions reproducibly detected in all three replicate spike samples of 12 fmol and above. The intensity progression of these ions with their corresponding spike level (vertical lines) can be visualized in Figure 3.6a for the second replicate of the spike series. Confirmation of the identity of the suspected spiked ions was obtained by coupling the nanoLC-chip device to an ion trap instrument. Examples of MS-MS spectra acquired using this targeted approach are presented in Figures 3.6b and 3.6c for m/z 480.6 and 435.2 (circled in Figure 3.6a), respectively. MS-MS spectra for the doubly protonated ion m/z 480.6 (Figure 3.6b) and m/z 435.2 (Figure 3.6c) enabled the unambiguous identification of peptides DTIVNELR and FAAYLER corresponding to

spiked proteins horseradish peroxidase and phosphorylase *b*, respectively. It is noteworthy that differential expression for spiked amount as low as 4.8 fmol were reproducibly detected.

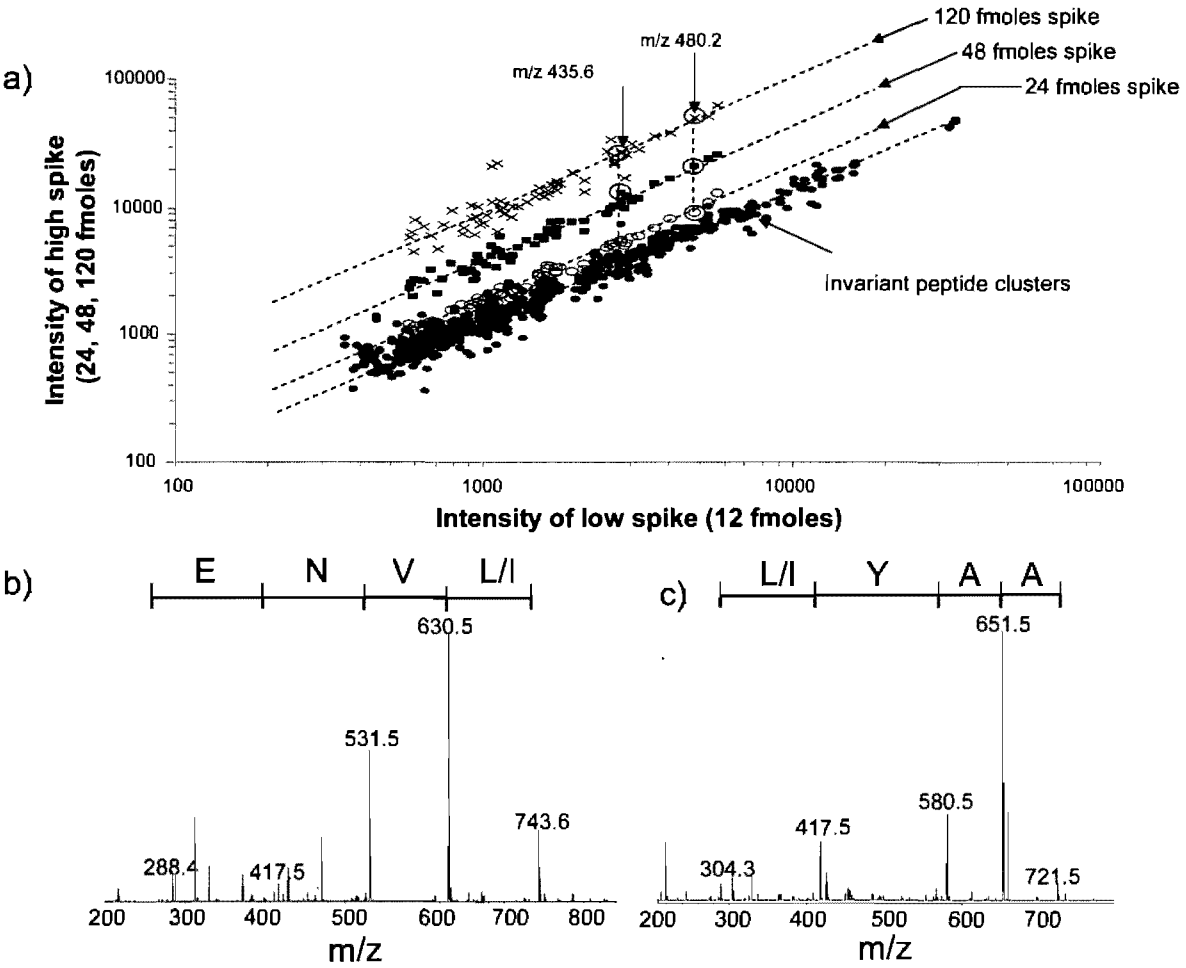

Figure 3.6: Detection of peptide ions from a digest of rat plasma proteins spiked with increasing amounts of an 8-protein tryptic digest. (a) Log-log plot of peptide intensity for protein spike levels of 24, 48 and 120 fmoles compared to 12 fmoles. Tandem mass spectra of (b) m/z 480.6 (horseradish peroxidase, DTIVNELR) and (c) m/z 435.2 (phosphorylase *b*, FAAYLER), both precursor ions are indicated by the arrows on Figure 3.6a.

Table 3.2 shows the linearity of response and limit of detection for 10 randomly selected spiked peptides. The limits of detection (S/N > 3) obtained for these ions ranged between 2.4 and 11.9 fmol. Excellent linearity of response was obtained across two orders of magnitude in intensity as indicated by the correlation coefficients ranging between 0.9928 and 0.9998. In the present situation, the dynamic range of peptide detection is approximately 14 parts per million given a loading capacity of 200 ng and a lower detection

boundary of 2 fmol for an average tryptic peptide molecular mass of 1350 Da. It is noteworthy that the present chip configuration provides a loading capacity of ~700 ng, which can be further extended by an order of magnitude using on-line two-dimensional chromatography.

Table 3.2: Linearity and limits of detection for a sample of spiked tryptic peptides reproducibly detected in rat plasma protein digest.

m/z	LOD (fmol)	S/N	r^2	Assigned proteins*
440.21 $^{2+}$	4.8	4.0	0.9983	Unassigned
453.25 $^{3+}$	11.9	6.6	0.9971	Phospho *b*
479.28 $^{1+}$	4.8	4.2	0.9960	Unassigned
501.76 $^{2+}$	11.9	3.5	0.9993	BSA
548.94 $^{3+}$	4.8	3.6	0.9983	G3PDH
566.21 $^{2+}$	4.8	3.7	0.9986	G3PDH
577.28 $^{2+}$	4.8	4.1	0.9990	BSA
593.28 $^{2+}$	2.4	3.4	0.9928	HSP
679.26 $^{2+}$	4.8	4.1	0.9998	ADH
721.84 $^{2+}$	4.8	4.5	0.9992	Phospho *b*

* Phospho *b*: Phosphorylase *b*, BSA: Bovine serum albumin, ADH: Alcohol dehydrogenase, G3PDH: Glyceraldehyde 3P dehydrogenase, HSP: Horse radish peroxidase

3.4.3 Application of the nanoLC-chip device with on-line two-dimensional chromatography

The modular design of the microfluidic LC system enables the integration of on-line two-dimensional chromatography (2D-LC) by inserting a SCX trapping column directly into the sample inlet port of the switching valve. In addition to providing enhanced selectivity, the coupling of 2D-LC also increases the overall number of detected peptides through higher sample loading. In the present investigation, we compared the overall improvements in sequence coverage and in the number of protein assigned for the nanoLC-chip system with and without an SCX trapping column. This comparison was performed on a biomarker sample obtained from rat plasma.

The analysis of serum samples typifies one the most significant analytical challenge in proteomics research given the diversity of proteins, their corresponding dynamic range in abundance (> 10 orders of magnitude), and the high proportion of proteins such as albumin and immunoglobulins [33, 34]. Indeed, 22 proteins alone represent more than 99 % of the plasma protein content [35]. Although common methods remove major proteins like albumin, IgG and transferrins, the proportion of these proteins still remain significant even after depletion. The inherent advantages of 2D-LC in terms of its high resolving power and increased loading capacity compare to one-dimensional reverse phase chromatography thus offers a means to address the overwhelming complexity and diversity of the plasma proteome [33, 36].

The number and distribution of major blood proteins obtained for the injection of 200 ng of an albumin and IgG-depleted rat plasma sample is shown in Figure 3.7a for data acquired with nanoLC-chip system coupled to an ion trap instrument. A total of 270 MS-MS spectra of good quality were acquired from a single nanoLC-chip-MS/MS run of which 206 were assigned to 47 non-redundant protein clusters. The largest proportion of the assigned spectra were matched to classic plasma proteins (36%) such as albumin, transferrin, complement components, and apolipoproteins. Cellular leakage proteins represented 17% whereas secreted or extracellular proteins corresponded to 18% of all proteins detected.

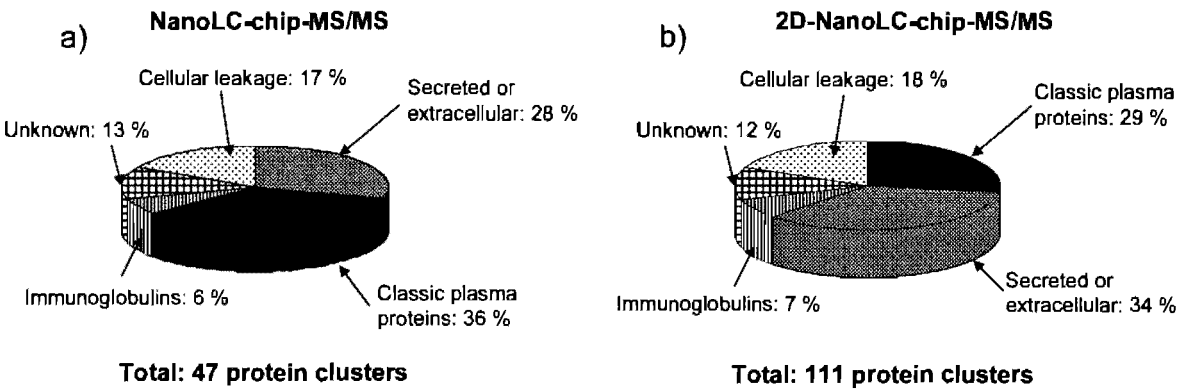

Figure 3.7: Distribution of the plasma proteins identified for the nanoLC-chip-MS/MS using (a) C_{18} reverse-phase and (b) two-dimensional (SCX/C_{18}) chromatography. A total of 200 ng and 1.5 µg was injected on the nanoLC-chip-MS/MS system for (a) and (b), respectively.

The same sample was subsequently analyzed using an on-line 2D-nanoLC-chip system. In this case, 1.5 μg of rat plasma digest was loaded on the SCX trap and eight fractions of different salt concentrations were injected to obtain a relatively uniform distribution of peptide ions with minimal carryover across fractions. A total of 2189 MS/MS spectra of good quality were obtained of which 1098 were assigned to 111 distinct protein clusters. Assigned proteins were distributed across secreted or extracellular (34%), classic plasma (29%) and cell leakage (18%) proteins (Figure 3.7b). Interestingly, classic proteins and immunoglobulins represented respectively less than 36% and 7% irrespective of the method used.

Table 3.3: Concentration and sequence coverage of proteins identified from rat plasma sample using nanoLC-chip-MS with and without two-dimensional chromatography

			Sequence coverage (%)	
Proteins	Mass (kDa)	Abundance in plasma Log conc. pg/mL*	LC-MS	2D-LC-MS
Hemopexin	51.3	8.7-9 [39]	17.0	40.9
Apolipoprotein A1	30.1	7-7.3 [34]	16.2	53.2
Complement C3	186.5	6.8-7.0 [34]	2.5	4.7
Haptoglobin	42.5	6.2-7 [34]	16.0	29.7
Transthyretin	15.8	5.9-6.2 [34]	0	10.2
Plasminogen	90.5	5.6-6.1 [34]	0	5.6
Retinol binding protein	23.2	5.5-5.8 [34]	0	5.0
insulin-like growth factor binding protein	67.0	4.7-5.0 [37]	0	5.1

* References are given in brackets

Furthermore, 2D-LC provided at least two to three times more sequence coverage (refers to the percentage of protein sequence having been identified) when compared to reverse-phase LC alone (Table 3.3). In the case of proteins such as plasminogen, insulin-like growth factor binding protein, retinol binding protein, and transthyretin, their identifications could only be obtained using 2D-LC. An example of this is shown in Figure 3.8 for the identification of insulin-like growth factor binding protein 1, a protein present at

80 ng/mL in plasma [37]. Insulin-like growth factor binding protein prolongs the half-life of the insulin growth factor (IGF) and can modulate growth factor-like activities on cell culture. This protein can also alter the interaction of IGFs with their cell surface receptors [38].

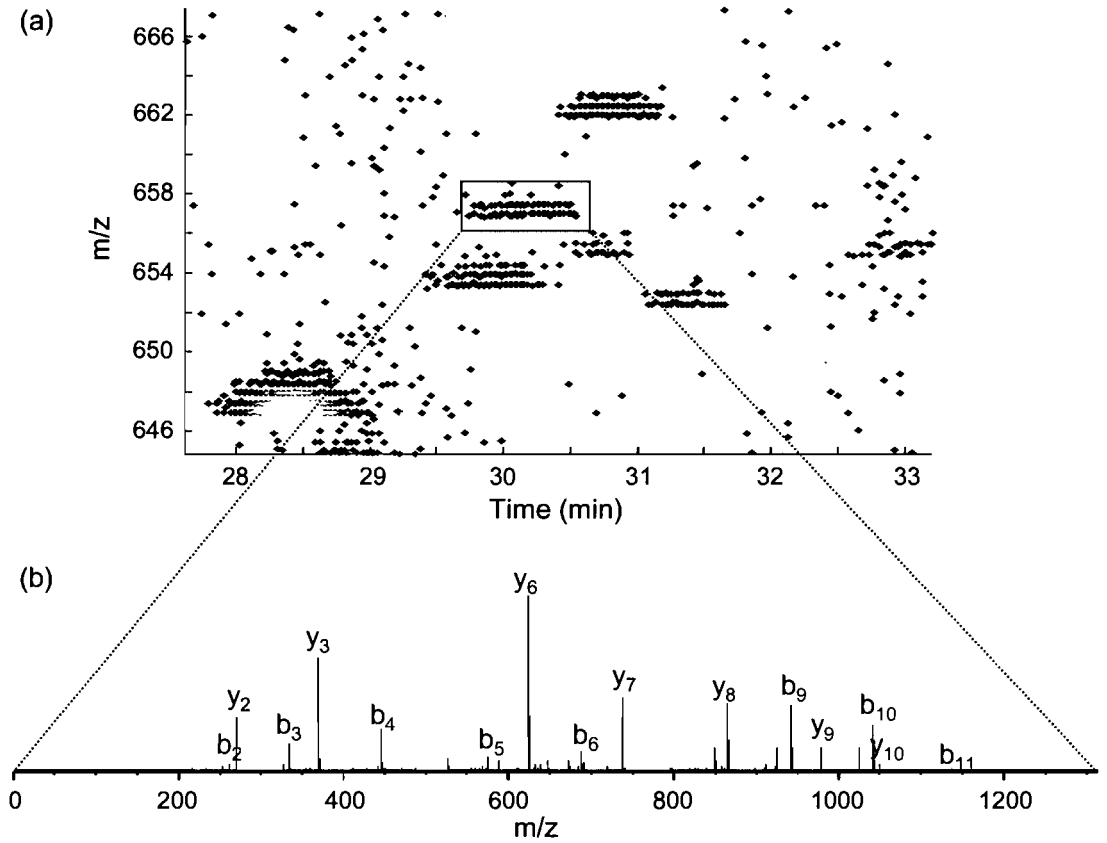

Figure 3.8: Identification of low-abundance plasma proteins from nanoLC-chip-MS/MS with two-dimensional (SCX/C_{18}) chromatography. (a) Narrow region of the contour profile of the 55 mM ammonium acetate fraction showing the doubly protonated peptide ion at m/z 657.0 (b) Tandem mass spectrum of m/z 657.0 corresponding to the tryptic peptide DFALQNPGVVPR from insulin-like growth factor binding protein 1.

3.5 Conclusion

The compact footprint of the present microfluidic system combined with its modularity and ease of operation provides significant advantages for the analysis of complex protein digests as part of proteomics discovery programs. Although endurance tests were not performed on this system, the chip device was operated in an uninterrupted manner for more than one hundred injections of complex tryptic digests with no apparent loss of performance. The figures of merit of this system in terms of reproducibility in retention time, m/z and intensity measurements were determined using tryptic digests comprising more than one thousand peptide ions. Peak measurements taken on more than 626 different peptide clusters from replicate injections (n=10) of an 8-protein digest provided RSD values of less than 0.5%, 0.003% and 9.1% for retention time, m/z, and intensity, respectively. In spite of its small dimensions, this microfluidic device provided peak capacity of 141 (50% height) for peptide elution taking place over 45 min. The integration of the precolumn, analytical column and nanoelectrospray emitter in a compact format also minimized dead volumes as reflected by the small peak width (0.3 min at 50% height) and asymmetry (< 1.2 at 10% height) measurements.

The excellent reproducibility of this microfluidic system facilitated the monitoring of changes in peptide abundance across complex and dense tryptic digests. In control experiments, 2-fold changes in differential abundance were detected for as low as 2.4 fmol of protein digests spiked in 100 ng plasma biomarker samples. The integration of SCX trapping column provides an increase in sample loading and selectivity thus facilitating the detection of low-abundance proteins such as that of insulin-like growth factor binding protein 1 present at less than 80 ng/mL in plasma biomarker samples. The ability to mine complex data sets with high sensitivity and reproducibility using microfluidic devices will be of great significance for protein expression and identification platforms given the expanding role of mass spectrometry-based proteomics in biomarker and drug discovery programs.

3.6 Acknowlegments

We thank Agilent Technologies for providing early access to the prototype LC-chip-MS system, MSD TOF and MSD trap XCT mass spectrometers. The authors also thank Agilent staff members Dr. John Michnowicz, and Georges Gauthier for their invaluable help and assistance, and Linda Cote and Sebastien Marchand for technical support. Finally, we acknowledge Caprion colleagues Eric Thibodeau for providing the depleted plasma sample, Alain Carrier and Anik Forest for providing the SCX column and Navdeep Jaitly for invaluable assistance with the clustering algorithms.

3.7 References

1. Domon, B. and Broder, S., *Implications of new proteomics strategies for biology and medicine.* J Proteome Res, 2004. **3**(2): p. 253-60.

2. Marko-Varga, G. and Fehniger, T.E., *Proteomics and disease--the challenges for technology and discovery.* J Proteome Res, 2004. **3**(2): p. 167-78.

3. Fountoulakis, M., Tsangaris, G., Oh, J.E., Maris, A., and Lubec, G., *Protein profile of the HeLa cell line.* J Chromatogr A, 2004. **1038**(1-2): p. 247-65.

4. Jacobs, J.M., Mottaz, H.M., Yu, L.R., Anderson, D.J., Moore, R.J., Chen, W.N., Auberry, K.J., Strittmatter, E.F., Monroe, M.E., Thrall, B.D., Camp, D.G., 2nd, and Smith, R.D., *Multidimensional proteome analysis of human mammary epithelial cells.* J Proteome Res, 2004. **3**(1): p. 68-75.

5. Schirle, M., Heurtier, M.A., and Kuster, B., *Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry.* Mol Cell Proteomics, 2003. **2**(12): p. 1297-305.

6. Zhu, W., Reich, C.I., Olsen, G.J., Giometti, C.S., and Yates, J.R., 3rd, *Shotgun proteomics of Methanococcus jannaschii and insights into methanogenesis.* J Proteome Res, 2004. **3**(3): p. 538-48.

7. Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T.G., Foss, E., Mao, Y., and Emili, A., *Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry.* Mol Cell Proteomics, 2004. **3**(10): p. 984-97.

8. Righetti, P.G., Campostrini, N., Pascali, J., Hamdan, M., and Astner, H., *Quantitative proteomics: a review of different methodologies.* Eur J Mass Spectrom (Chichester, Eng), 2004. **10**(3): p. 335-48.

9. Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T.A., Hill, L.R., Norton, S., Kumar, P., Anderle, M., and Becker, C.H., *Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards.* Anal Chem, 2003. **75**(18): p. 4818-26.

10. Tessier, S., Bonneil, E., Carrier, A., and Thibault, P. in *Proc. 52nd ASMS Conference on Mass Spectrometry and Allied Topics.* 2004. Nashville, TN.

11. Shen, Y., Moore, R.J., Zhao, R., Blonder, J., Auberry, D.L., Masselon, C., Pasa-Tolic, L., Hixson, K.K., Auberry, K.J., and Smith, R.D., *High-efficiency on-line*

solid-phase extraction coupling to 15-150-microm-i.d. column liquid chromatography for proteomic analysis. Anal Chem, 2003. **75**(14): p. 3596-3605.

12. Shen, Y., Zhao, R., Berger, S.J., Anderson, G.A., Rodriguez, N., and Smith, R.D., *High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics.* Anal Chem, 2002. **74**(16): p. 4235-49.

13. Lion, N., Rohner, T.C., Dayon, L., Arnaud, I.L., Damoc, E., Youhnovski, N., Wu, Z.Y., Roussel, C., Josserand, J., Jensen, H., Rossier, J.S., Przybylski, M., and Girault, H.H., *Microfluidic systems in proteomics.* Electrophoresis, 2003. **24**(21): p. 3533-62.

14. Marko-Varga, G., Nilsson, J., and Laurell, T., *New directions of miniaturization within the proteomics research area.* Electrophoresis, 2003. **24**(21): p. 3521-32.

15. Belder, D. and Ludwig, M., *Surface modification in microchip electrophoresis.* Electrophoresis, 2003. **24**(21): p. 3595-606.

16. Bandilla, D. and Skinner, C.D., *Capillary electrochromatography of peptides and proteins.* J Chromatogr A, 2004. **1044**(1-2): p. 113-29.

17. Deng, Y., Zhang, H., and Henion, J., *Chip-based quantitative capillary electrophoresis/mass spectrometry determination of drugs in human plasma.* Anal Chem, 2001. **73**(7): p. 1432-9.

18. Figeys, D. and Aebersold, R., *Nanoflow solvent gradient delivery from a microfabricated device for protein identifications by electrospray ionization mass spectrometry.* Anal Chem, 1998. **70**(18): p. 3721-7.

19. Lazar, I.M., Ramsey, R.S., Jacobson, S.C., Foote, R.S., and Ramsey, J.M., *Novel microfabricated device for electrokinetically induced pressure flow and electrospray ionization mass spectrometry.* J Chromatogr A, 2000. **892**(1-2): p. 195-201.

20. Li, J., Thibault, P., Bings, N.H., Skinner, C.D., Wang, C., Colyer, C., and Harrison, J., *Integration of microfabricated devices to capillary electrophoresis-electrospray mass spectrometry using a low dead volume connection: application to rapid analyses of proteolytic digests.* Anal Chem, 1999. **71**(15): p. 3036-45.

21. Xue, Q., Foret, F., Dunayevskiy, Y.M., Zavracky, P.M., McGruer, N.E., and Karger, B.L., *Multichannel microchip electrospray mass spectrometry.* Anal Chem, 1997. **69**(3): p. 426-30.

22. Li, J., LeRiche, T., Tremblay, T.L., Wang, C., Bonneil, E., Harrison, D.J., and Thibault, P., *Application of microfluidic devices to proteomics research: identification of trace-level protein digests and affinity capture of target peptides.* Mol Cell Proteomics, 2002. **1**(2): p. 157-68.

23. Li, J., Wang, C., Kelly, J.F., Harrison, D.J., and Thibault, P., *Rapid and sensitive separation of trace level protein digests using microfabricated devices coupled to a quadrupole--time-of-flight mass spectrometer.* Electrophoresis, 2000. **21**(1): p. 198-210.

24. Xue, Q., Dunayevskiy, Y.M., Foret, F., and Karger, B.L., *Integrated multichannel microchip electrospray ionization mass spectrometry: analysis of peptides from on-chip tryptic digestion of melittin.* Rapid Commun Mass Spectrom, 1997. **11**(12): p. 1253-6.

25. Kutter, J.P., Jacobson, S.C., Matsubara, N., and Ramsey, J.M., *Solvent-programmed microchip open-channel electrochromatography.* Anal Chem, 1998. **70**: p. 3291-3297.

26. Kutter, J.P., Jacobson, S.C., and Ramsey, J.M., *Solid phase extraction on microfuidic devices.* J Microcolum Sep, 2000. **12**: p. 93-97.

27. Yu, C., Davey, M.H., Svec, F., and Frechet, J.M., *Monolithic porous polymer for on-chip solid-phase extraction and preconcentration prepared by photoinitiated in situ polymerization within a microfluidic device.* Anal Chem, 2001. **73**(21): p. 5088-96.

28. Yu, C., Xu, M., Svec, F., and Frechet, J.M., *Preparation of monolithic polymers with controlled porous properties for microfluidic chip applications using photoinitiated free-radical polymerization.* J Polym Sci Polym Chem, 2002. **40**: p. 755-769.

29. Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., and van de Goor, T., *Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip.* Anal Chem, 2005. **77**(2): p. 527-33.

30. Kearney, P. and Thibault, P., *Bioinformatics meets proteomics--bridging the gap between mass spectrometry data analysis and cell biology.* J Bioinform Comput Biol, 2003. **1**(1): p. 183-200.

31. Giddings, J.C., *Maximum number of components resolvable by gel filtration and other elution chromatographic methods.* Anal Chem, 1967. **39**(8): p. 1027-1028.

32. Lan, K. and Jorgenson, J.W., *Automated measurement of peak widths for the determination of peak capacity in complex chromatograms.* Anal Chem, 1999. **71**(3): p. 709-14.

33. Adkins, J.N., Varnum, S.M., Auberry, K.J., Moore, R.J., Angell, N.H., Smith, R.D., Springer, D.L., and Pounds, J.G., *Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry.* Mol Cell Proteomics, 2002. **1**(12): p. 947-55.

34. Anderson, N.L. and Anderson, N.G., *The human plasma proteome: history, character, and diagnostic prospects.* Mol Cell Proteomics, 2002. **1**(11): p. 845-67.

35. Tirumalai, R.S., Chan, K.C., Prieto, D.A., Issaq, H.J., Conrads, T.P., and Veenstra, T.D., *Characterization of the low molecular weight human serum proteome.* Mol Cell Proteomics, 2003. **2**(10): p. 1096-103.

36. Shen, Y., Jacobs, J.M., Camp, D.G., 2nd, Fang, R., Moore, R.J., Smith, R.D., Xiao, W., Davis, R.W., and Tompkins, R.G., *Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome.* Anal Chem, 2004. **76**(4): p. 1134-44.

37. Heald, A.H., Siddals, K.W., Fraser, W., Taylor, W., Kaushal, K., Morris, J., Young, R.J., White, A., and Gibson, J.M., *Low circulating levels of insulin-like growth factor binding protein-1 (IGFBP-1) are closely associated with the presence of macrovascular disease and hypertension in type 2 diabetes.* Diabetes, 2002. **51**(8): p. 2629-36.

38. Anderson, N.L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R.S., Conrads, T.P., Veenstra, T.D., Adkins, J.N., Pounds, J.G., Fagan, R., and Lobley, A., *The human plasma proteome: a nonredundant list developed by combination of four separate sources.* Mol Cell Proteomics, 2004. **3**(4): p. 311-26.

39. Anderson, N.L., Anderson, N.G., Haines, L.R., Hardie, D.B., Olafson, R.W., and Pearson, T.W., *Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA).* J Proteome Res, 2004. **3**(2): p. 235-44.

4. The MHC class I peptide repertoire is molded by the transcriptome

Marie-Hélène Fortier, Etienne Caron, Marie-Pierre Hardy,
Gregory Voisin, Claude Perreault, and Pierre Thibault

The Journal of Experimental Medicine, 2008. 205(3): p. 595-610

4.1 Abstract

Under steady-state conditions, major histocompatibility complex (MHC) I molecules are associated with self-peptides that are collectively referred to as the MHC class I peptide (MIP) repertoire. Very little is known on the genesis and molecular composition of the MIP repertoire. We developed a novel high-throughput mass spectrometry approach that yields an accurate definition of the nature and relative abundance of unlabeled peptides presented by MHC I molecules. We identified 189 and 196 MHC I-associated peptides from normal and neoplastic mouse thymocytes, respectively. By integrating our peptidomic data with global profiling of the transcriptome, we reached two conclusions. The MIP repertoire of primary mouse thymocytes is biased toward peptides derived from highly abundant transcripts and is enriched in peptides derived from cyclins/cyclin-dependent kinases and helicases. Furthermore, we found that ~25% of MHC I-associated peptides were differentially expressed on normal versus neoplastic thymocytes. Approximately half of those peptides derived from molecules directly implicated in neoplastic transformation (e.g., components of the PI3K-AKT-mTOR pathway). In most cases, overexpression of MHC I peptides on cancer cells entailed posttranscriptional mechanisms. Our results show that high-throughput analysis and sequencing of MHC I-associated peptides yields unique insights into the genesis of the MIP repertoire in normal and neoplastic cells.

4.2 Introduction

MHC class I molecules present short peptides at the cell surface, typically 8-11 mers, for scrutiny by CD8 T lymphocytes[1, 2]. Generation of peptide-MHC I complexes is initiated by proteasomal degradation of source proteins in the cytosol [3]. Peptides generated by the proteasome are translocated in the endoplasmic reticulum where they are subjected to N-terminal trimming, and then incorporated in MHC I proteins and exported at the cell surface [4-7]. Presentation of microbial peptides by MHC I is required to elicit CD8 T cell responses against intracellular pathogens [8]. Under steady-state conditions, i.e., in the absence of infection, cell surface MHC I molecules are associated solely with self-peptides. These peptides, collectively referred to as the MHC class I peptide (MIP) repertoire [9], play vital roles. They shape the repertoire of developing thymocytes [10, 11], transmit survival signals to mature CD8 T cells [12], amplify responses against intracellular pathogens [13], allow immunosurveillance of neoplastic cells [14], and influence mating preferences in mice [15]. The MIP repertoire is also involved in immunopathology because it can be targeted by autoreactive T cells that initiate autoimmune diseases and alloreactive T cells that cause graft rejection and graft-versus-host disease [16, 17].

Despite the tremendous importance of the MIP repertoire, we know very little about its genesis and molecular composition. Proteomic analysis of the MIP repertoire is a daunting task because estimates suggest that it encompasses thousands of peptides present in low copy numbers per cell [1, 18]. Each MHC molecule recognizes peptides through a broadly defined consensus motif of amino acids serving as anchors to the appropriate binding pockets on the MHC molecules. Such motifs were first established by pool Edman sequencing of unfractionated peptide mixtures eluted from MHC molecules [19]. Direct biochemical characterization of specific MHC I-associated peptides has typically involved immunoaffinity purification of MHC molecules after cell lysis, fractionation of the peptides by chromatography, and sequencing, initially by Edman's method and, more recently by mass spectrometry (MS). Refinements in MS methods pioneered by Hunt *et al.* represented major progress and led to characterization of several MHC I-associated peptides [18, 20, 21] in spite of limitations such as low peptide yield, preferential loss of peptides with low

affinity for MHC I, and contamination of the MHC molecules by cellular debris and detergents [22, 23].

Two high-throughput strategies were therefore implemented to provide comprehensive molecular definition of the MIP repertoire. The first is based on transfection of cell lines with expression vectors coding soluble secreted MHCs (lacking a functional transmembrane domain) and elution of peptides associated with secreted MHCs [24, 25]. This interesting approach improves MHC I peptide recovery, but presents some limitations, as follows: (a) it cannot be used on freshly explanted cells; (b) cell transfection by itself may perturb the MIP repertoire [26]; (c) the MIP repertoire associated with soluble MHC corresponds to the repertoire of peptides that can bind the transfected MHC allele (what "can be presented") but not necessarily to peptides that are normally presented at the cell surface (what "is presented"). The second approach hinges on chemical or metabolic labeling to provide quantitative profiles of MHC I-associated peptides [9, 27-29]. Although chemical derivatization suffers from variable modification yields and unexpected side reaction products, metabolic labeling is only applicable to certain MHC I allelic products and to cell culture model systems [30, 31]. Nonetheless, high-throughput peptide sequencing analyses have provided crucial insights into the structure of the MIP repertoire. First, the source proteins for the MIP repertoire are found in almost every compartment of the cell [25]. Second, only a limited correlation is observed between the amounts of the MHC I-associated peptides presented by cells and the relative expression of source proteins from which these peptides are derived [28]. A likely explanation for this discrepancy is that the MIP repertoire preferentially derives from defective ribosomal products (DRiPs) and short-lived proteins relative to slowly degraded proteins [2, 32, 33]. Third, for peptides differentially expressed on normal versus neoplastic cells, no clear correlation was found between mRNA levels and corresponding MHC peptide levels [9].

The goal of our work was to understand the structure and genesis of the MIP repertoire. In accordance with a recent advocacy for "systems immunology" [34], we used novel bioinformatics tools based on peptide mapping and segmentation to obtain a global and accurate quantification of native unlabeled peptides present in the MIP repertoire (unpublished data)[35, 36]. In particular, we addressed four specific questions. Does the MIP repertoire reflect the composition of the transcriptome? Are some gene families

overrepresented in the MIP repertoire? What is the impact of neoplastic transformation on the MIP repertoire? Can a high-throughput, label-free peptide sequencing platform provide valuable insights for the identification of immunogenic tumor-associated epitopes?

4.3 Results

4.3.1 Experimental approach for the identification and quantification of MHC I-associated peptides

Mild acid elution (MAE) of MHC I-associated peptides from living cells presents several advantages over immunoprecipitation of peptide-MHC I complexes. Because it involves fewer purification steps and no detergents, MAE yields ~10 times more MHC I-associated peptides than the latter and introduces no bias linked to preferential loss of low-affinity peptides [37-39]. However, MAE has never been used for high-throughput sequencing of the MIP repertoire because it is assumed that eluted peptides contain not only MHC I-associated peptides but also "contaminant" peptides. We reasoned that this problem could be overcome by using MHC I-deficient cells as a negative control. Because β2-microglobulin (β_2m) is essential for formation of stable peptide-MHC I complexes, β_2m-deficient cells are MHC I-deficient. We therefore compared peptides eluted from three cell populations, as follows: WT EL4 thymoma cell line, a β_2m-deficient EL4 mutant cell line (C4.4-25$^-$), and C4.4-25$^-$ cells transfected with a genomic clone of murine β_2m (E50.16$^+$) [40]. Flow cytometry analysis showed that expression of H2D^b, H2K^b and Qa2 MHC I molecules were abrogated in β_2m-deficient EL4 cells but was restored in β_2m transfectants (Figure 4.1A).

Samples obtained by MAE of peptides from 8 x 10^7 cells were analyzed using an on-line 2D-nanoLC-MS system (see Materials and methods for more details). Each individual LC-MS run was visualized as a three-dimensional map where each isotopic profile (dot) corresponds to a given ion that is defined by its specific abundance, m/z and retention time coordinates (Figure 4.1B). We first assessed the reproducibility of data obtained with our method using peptide eluates from WT EL4 cells. We found that 95% of peptide ions showed a variation of less than ± 1.4-fold and ± 2.2-fold in abundance across 2D-nanoLC-MS instrumental and biological replicates, respectively (Figure 4.2). Therefore, peptides were considered to be differentially expressed when the fold difference in abundance was ≥ 2.5 ($P < 0.05$). We next compared profiles from WT and β_2m-deficient EL4 cells. Contour profiles obtained after computational analyses showed a higher level of peptide complexity

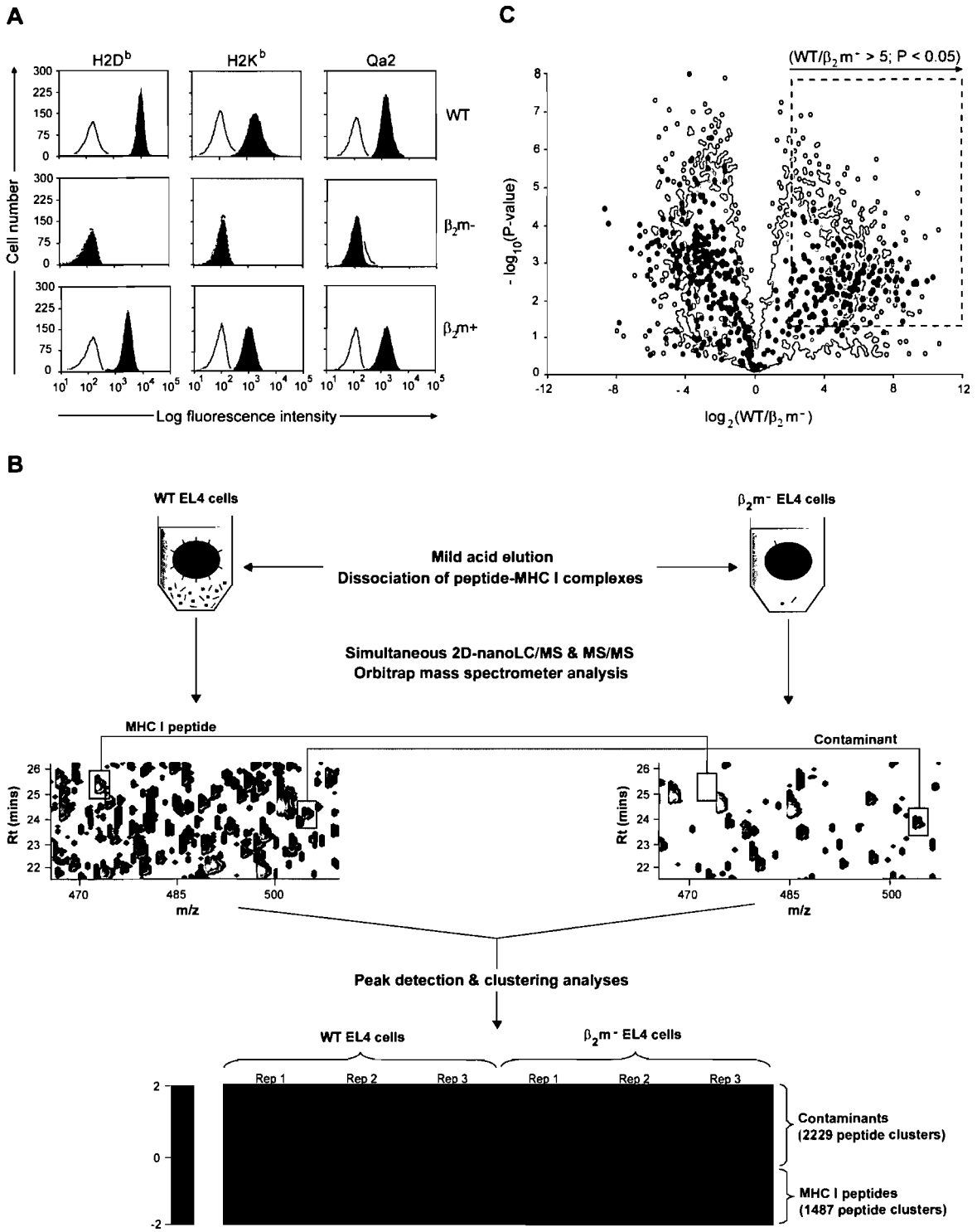

Figure 4.1: Experimental design for identification and relative quantification of native unlabeled MHC I-associated peptides. (A) Cell surface MHC I expression on EL4 WT, β_2m^- mutant and β_2m^+ transfectant cell lines. Cells were stained with antibodies against $H2D^b$, $H2K^b$ and Qa2 (black) or the corresponding isotype control antibody (gray). (B) Peptides obtained by MAE were analyzed by on-line 2D-nanoLC-MS (three biological replicates for each cell population). Contour profiles of m/z versus retention time

versus intensity were used to visualize differences between MS profiles (middle). A logarithmic intensity scale distinguishes between low (dark red) and highly (bright yellow) abundant species. Examples of peptides that were differentially expressed (blue line) or not (green line) between WT and β_2m^- EL4 cells are highlighted in boxes. Heat map representation shows differential peptide expression between WT and β_2m^- EL4 cells where each horizontal line corresponds to a unique peptide cluster (n = 3,716) (bottom). A logarithmic scale depicts peptides that are expressed at high (red) or low (green) level in each cell population. (C) Volcano plot representation showing reproducibly detected peptides ions across three replicate analyses. Peptide clusters (n = 1,236) highlighted in dashed box were considered as class I-restricted ($P \leq 0.05$ and fold change ≥ 5). Peptides that were sequenced by MS/MS are represented by colored dots: green for contaminants, and blue for MHC I-associated peptides.

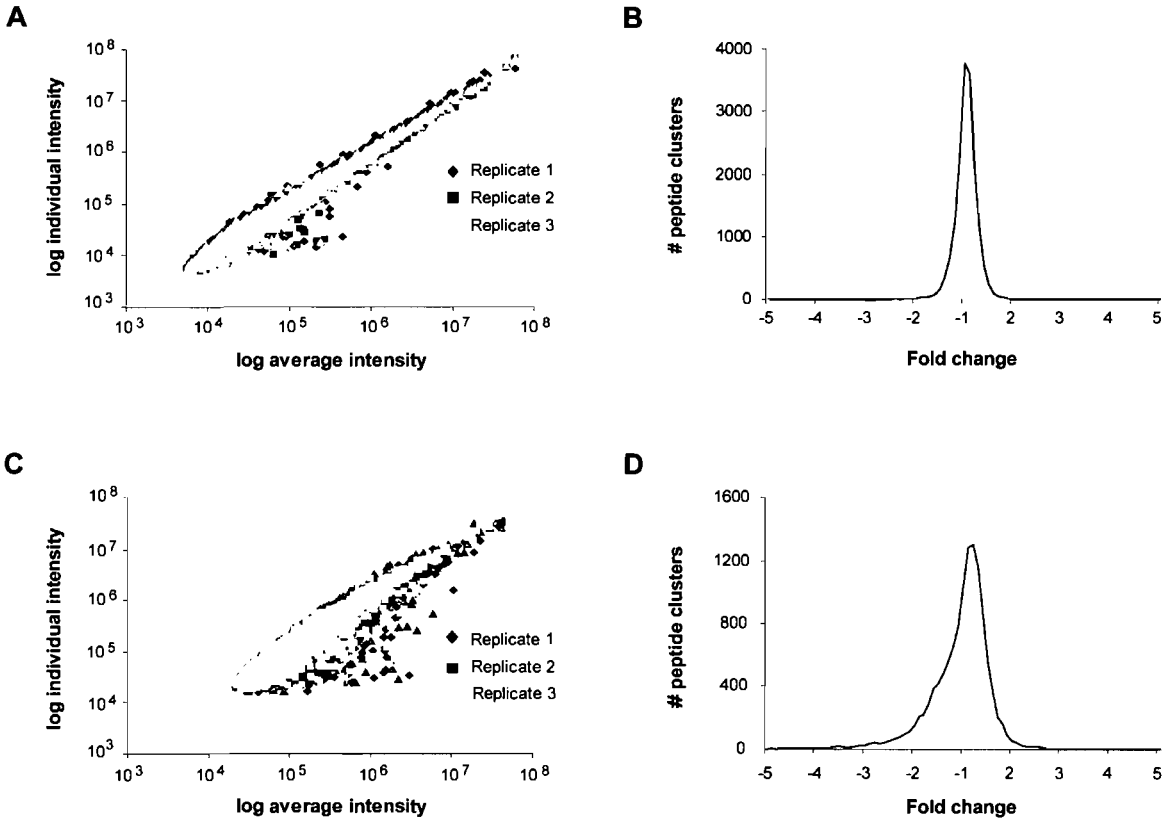

Figure 4.2: Reproducibility of intensity measurements for instrumental and biological replicates. (A) Log-log plot of individual ion intensity versus the average intensity for replicate 2D-LC analysis (n =3). (B) Expression plot for 2D-LC instrumental replicates (n = 5,568 peptide clusters). Distribution indicates that 95% of peptide ions show a variation of less than 1.4-fold change in intensity. (C) Scatter plot of intensities for biological replicates (n =3). (D) Expression plot from different biological preparations (n =4,046 peptide clusters). Peptide ions (95%) show variation in intensity within +/– 2.2-fold change.

for WT compared to β_2m-deficient EL4 cells (Figure 4.1B). We assessed the proportion of MHC I-associated versus contaminant peptides using in-house peptide detection software (unpublished data). To identify and define individual peptide clusters, peptide lists generated by Mass Sense were compared through replicate injections from both WT- and β_2m^- mutant-derived samples using segmentation analyses with hierarchical clustering. Heat map representation was used to visualize differences in peptide cluster expression between WT- and β_2m^- mutant-derived samples (Figure 4.1B). Peptides that were recovered only from WT EL4 cells were considered to be part of the MIP repertoire. Because β_2m-deficient cells express low amounts of incomplete MHC I molecules at the cell surface [41], we also considered that peptides significantly more abundant on WT relative to β_2m-deficient cells were MHC I-associated. In the latter case, we used a very stringent criterion; to be considered as MHC I-associated, a peptide had to be at least 5 times more abundant on WT relative to β_2m-deficient cells. Out of 3,716 unique peptide clusters that were reproducibly detected across three biological replicates, 1,487 peptide clusters were overexpressed or uniquely detected within the WT EL4 sample. Thus, we estimated that ~40% of acid-eluted materials correspond to specific-MHC I peptides.

We obtained MS/MS assignments for 881 of the 3,716 peptide clusters, all of which had a characteristic fragmentation associated to peptide backbone cleavages. Although these identifications correspond to 24% of the entire ion population detected, other MS/MS spectra were assigned to modified residues, de novo peptide sequences or spectra of poor quality. Peptide coordinates were computationally related to their corresponding sequenced peptides. Out of 881 sequenced peptides, 383 had a consensus motif for $H2D^b$, $H2K^b$, Qa1 or Qa2 MHC I molecules. Stringent validation criteria were next applied to select peptide ions showing mass accuracy of < 30 parts per million (ppm), with MS/MS spectra displaying consistent sets of b- and/or y-type fragment ion series corresponding to the MHC I peptide sequence. This manual validation enabled the identification of 178 unique MS/MS corresponding to MHC I-associated peptides: 158 were detected uniquely in eluates from WT EL4 cells, while 20 were overexpressed at least 5-fold on WT relative to β_2m-deficient EL4 cells (Figure 4.1C). To confirm that differences between WT and β_2m-deficient EL4 cells were truly β_2m dependent, we analyzed peptides recovered from $E50.16^+$ cells (β_2m-deficient EL4 cells transfected with a genomic clone of murine β_2m).

Table 4.1: Relative abundance of a representative set of peptides recovered in eluates from three EL4 cell variants: WT, β_2m^- mutants and β_2m^+ transfectants.

Gene ID	Gene Symbol	Sequence	WT	β_2m^-	β_2m^+	WT/β_2m^-
20848	*Stat3*	ATLVFHNL	4,023,137	12,000	6,208,130	335
66165	*Bccip*	KAPVNTAEL	1,741,908	12,000	998,607	145
15016	*H2-Q5*	AMAPRTLLL	5,436,819	12,000	12,941,105	453
67755	*Ddx47*	KTFLFSATM	70,059	12,000	289,017	6
12649	*Chek1*	TGPSNVDKL	6,545,497	112,308	14,733,333	58
267019	*Rps15a*	VIVRFLTV	514,406	12,000	12,000	43
71745	*Cul2*	VINSFVHV	67,372	12,000	244,370	6
233489	*Picalm*	NGVINAAFM	265,420	12,000	438,191	22
26413	*Mapk1*	VGPRYTNL	39,166,667	149,969	45,832,060	261
108655	*Foxp1*	QQLQQQHLL	705,266	12,000	1,054,408	59
170755	*Sgk3*	YSIVNASVL	255,048	38,491	151,335	7
326622	*Upf2*	SAVIFRTL	244,247	12,000	1,102,686	20
12877	*Cpeb1*	SMLQNPLGNVL	185,198	12,000	528,268	15
94176	*Dock2*	SMVQNRVFL	895,081	12,000	671,603	75
16913	*Psmb8*	GGVVNMYHM	3,332,728	12,000	2,116,510	277

Gene ID and Gene Symbol description refer to NCBI gene entries. Intensities for each EL4 variant cell line correspond to the average MS signal calculated from triplicate experiments. An intensity threshold value of 12,000 was fixed when no signal was detected. To be considered as MHC I-associated, a peptide had to be at least 5 times more abundant on WT relative to β_2m-deficient cells.

As expected, we observed that 95% of MHC I peptides were expressed at the cell surface of β_2m^+ transfectants. Absence of a few MHC I peptides on β_2m transfectants is probably explained by the lower expression of $H2D^b$ and $H2K^b$ on β_2m transfectants compared with WT cells (Figure 4.1A). For the 178 sequenced MHC I-associated peptides, relative abundance in the three EL4 cell variants (WT, β_2m^- mutants and β_2m^+ transfectants) is presented in Table AII.I. Data for a representative set of peptides is shown in Table 4.1. Notably, although MHC I peptides were 8-13 mers, the length of contaminant peptides was more variable, ranging from 6 to 26 amino acids with a median value of 13 residues (unpublished data).

4.3.2 Definition of the MIP repertoire presented by discrete MHC I allelic products

Large databases of MHC I peptides and computational binding methods are now part of emerging biomedical resources [42-45]. Information in these databases concerns mainly, though not exclusively, viral and bacterial-derived peptides presented by MHC I allelic products in humans and mice. Each MHC I peptide in Table AII.I was manually classified according to restriction size and binding motif favored by D^b, K^b, Qa1 and Qa2 class I molecules [19, 46, 47]. From 178 MHC I peptides identified from variant EL4 cell lines, 92% were presented by classical MHC Ia allelic products $H2D^b$ and $H2K^b$, and 8% were presented by MHC Ib molecules Qa1 and Qa2 (Figure 4.3A, Table AII.II). When viewed *in toto*, peptide pools presented by $H2D^b$, $H2K^b$ and Qa2 displayed the canonical binding motifs documented for these MHC I molecules (Figure 4.3B).

A priori, two factors may determine whether specific peptides are presented by MHC I: peptide affinity for MHC molecules and the processing of source proteins along the MHC I presentation pathway. Protein attributes that are relevant to MHC I processing include rates of protein translation, DRiP formation and degradation [32, 33, 48, 49]. To evaluate the importance of peptide binding affinity, each of the 164 source proteins of peptides presented by $H2D^b$ (n = 96) or $H2K^b$ (n = 68) was analyzed with the smm algorithm [44]. We scored and ranked the predicted binding affinity of all peptides from individual proteins for $H2D^b$ or $H2K^b$ (for example, a protein of 400 amino acids contains 392 nonapeptide sequences). We then asked how each peptide that we had sequenced by MS/MS would rank

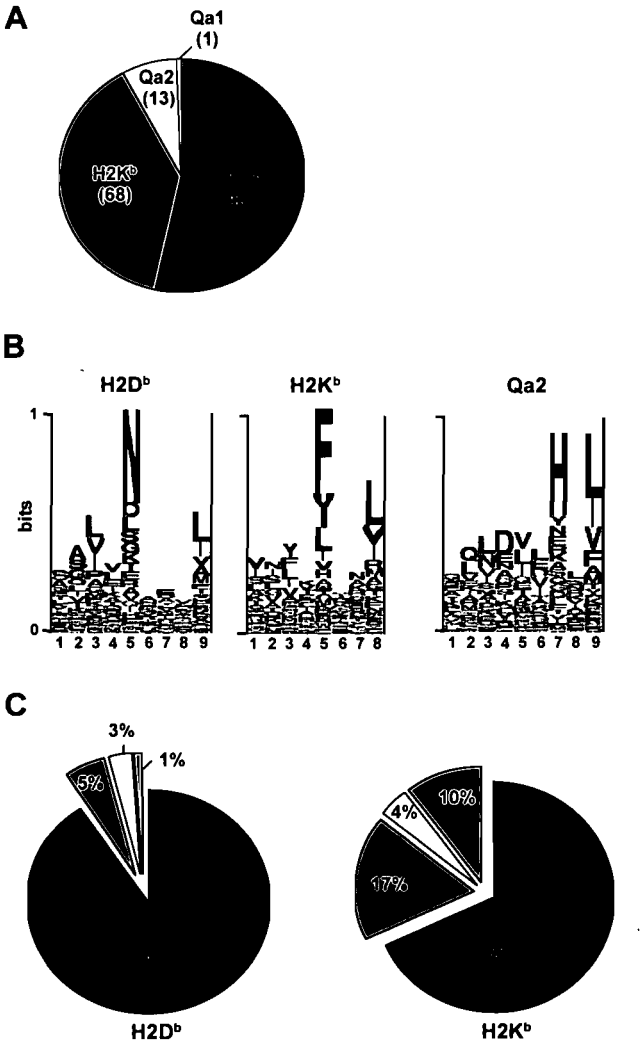

Figure 4.3: Allelic distribution and binding scores of 178 MHC I-associated peptides eluted from EL4 cells. (A) Pie chart shows distribution of 178 MHC I-associated peptides eluted from EL4 cells. The smm, SYFPEITHI ($H2D^b$ and $H2K^b$) and Rankpep (Qa2) computational models were used to link individual peptides to MHC I allelic products. (B) Logo showing the profile motif for peptides presented by $H2D^b$, $H2K^b$ and Qa2 molecules. Acidic (red), basic (blue), hydrophobic (black) and neutral (green) amino acids are illustrated. (C) Individual source proteins for peptides presented by $H2D^b$ and $H2K^b$ (n = 164) were entered in the smm binding algorithm. We assessed the predicted MHC I binding affinity of all peptides contained in individual proteins. Pie charts show the proportion of peptides eluted from EL4 cells that ranked in the top 1% (blue), top 5% (green), top 10% (yellow), or below the 90^{th} percentile of peptides (red).

relative to other peptides from its source protein. Remarkably, 91% of $H2D^b$-associated peptides eluted from EL4 cells ranked in the top 1% in terms of $H2D^b$ binding affinity (Figure 4.3C, Table AII.II). Similar results were obtained from SYFPEITHI database

(unpublished data). That some peptide sequences predicted to be top MHC I binders were not found in the MIP repertoire of EL4 cells is not surprising. This means either that proteases involved in MHC I processing do not generate these specific "degradation products" or that they are present in subthreshold amounts. For $H2K^b$, 85% of peptides ranked in the top 5% for predicted $H2K^b$ binding affinity (Figure 4.3C, Table AII.II). These data support the concept that MHC I binding affinity plays a dominant role in shaping the content of the MIP repertoire.

4.3.3 Discrimination between MHC I-associated peptides and contaminant peptides using bioinformatic tools

In the aforementioned experiments, we used β_2m-deficient cells as negative control to discriminate MHC I-associated peptides from contaminant peptides. The need for an MHC I-deficient negative control would nevertheless represent a significant hurdle for high-throughput sequencing of the MIP repertoire of primary cells.

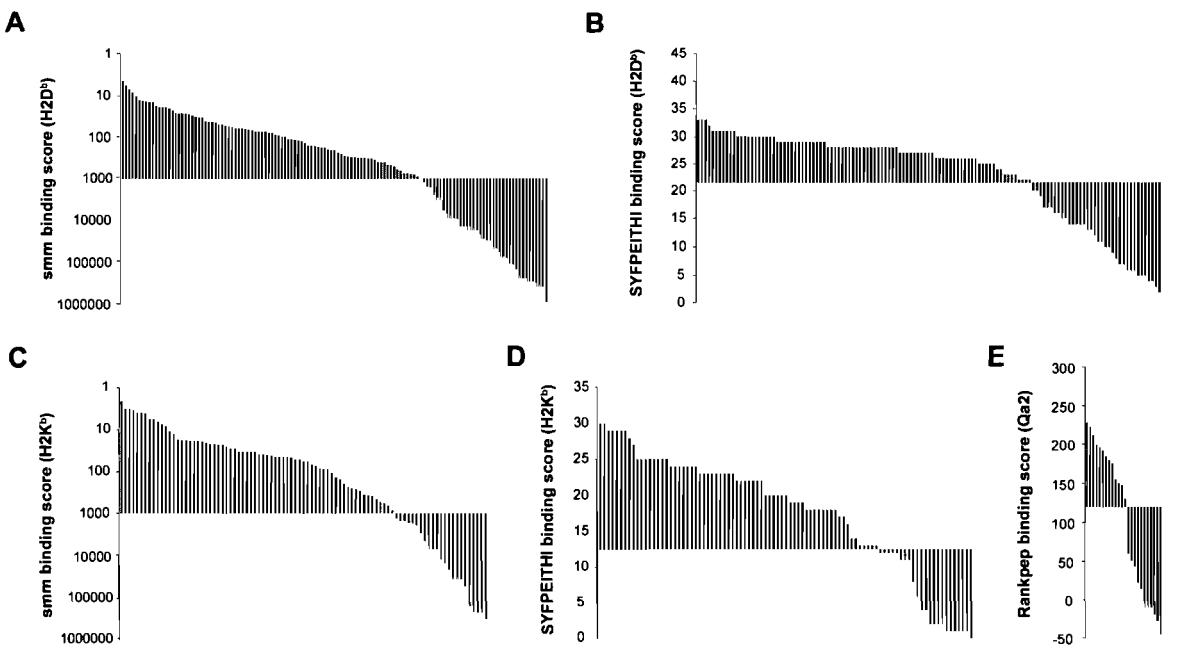

Figure 4.4: Discrimination between MHC I-associated peptides and contaminant peptides using bioinformatic tools. (A-E) For peptides eluted from EL4 cells, the y axis shows computed MHC binding scores determined with the smm (A, C), SYFPEITHI (B, D) and Rankpep (E) computational methods. The x axes cut at the selected binding thresholds. Each bar represents a sequenced peptide. Individual $H2D^b$-, $H2K^b$-, Qa2-associated peptides (red) and contaminant peptides (blue) were scored as illustrated.

We therefore asked whether bioinformatic tools could be used to identify MHC I-associated peptides among a peptide mixture obtained by MAE. We first ranked the 164 $H2D^b$- and $H2K^b$-associated peptides eluted from EL4 cells as a function of their MHC binding score estimated with smm and SYFPEITHI algorithms (Figure 4.4, red bars). We found that ~97% of MHC I-associated peptides were scored below a smm binding threshold of 1,000 nM (IC_{50}) and above of SYFPEITHI binding threshold of 22 for $H2D^b$ and 13 for $H2K^b$ (Figure 4.4). The aforementioned thresholds therefore provided a 3% false-negative rate (3% of MHC I-associated peptides were wrongly classified as contaminants). To estimate the false-positive rate obtained with these thresholds, we analyzed the 215 contaminant peptides eluted from EL4 cells representing high quality assignment (Mascot score >45, mass accuracy < 30 ppm, MS/MS manually inspected) from a list of 498 candidates that were not significantly overexpressed on WT relative to β_2m-deficient EL4 cells. Only 56 out of the 215 contaminant peptides had the canonical length for binding $H2D^b$ or $H2K^b$, and very few of them satisfied our smm and SYFPEITHI thresholds (Figure 4.4, blue bars). Overall, the false-positive rate was < 2% (< 2% of contaminants were wrongly classified as MHC I-associated peptides). Similar results were obtained for Qa2 peptides using the Rankpep algorithm with a binding threshold of 120. We synthesized nine peptides across the predicted affinity spectrum and tested their ability to bind $H2K^b$ or $H2D^b$. Peptides classified as MHC I-associated according to their smm score did bind to MHC I, whereas peptides classified as contaminants did not (Figure 4.5). These experimental data further validate bioinformatic discrimination between MHC I-associated peptides and contaminant peptides. We therefore conclude that peptides obtained by MAE can be categorized as MHC I-associated versus contaminant peptides with 1-3% false-positive and -negative rates using publicly available algorithms.

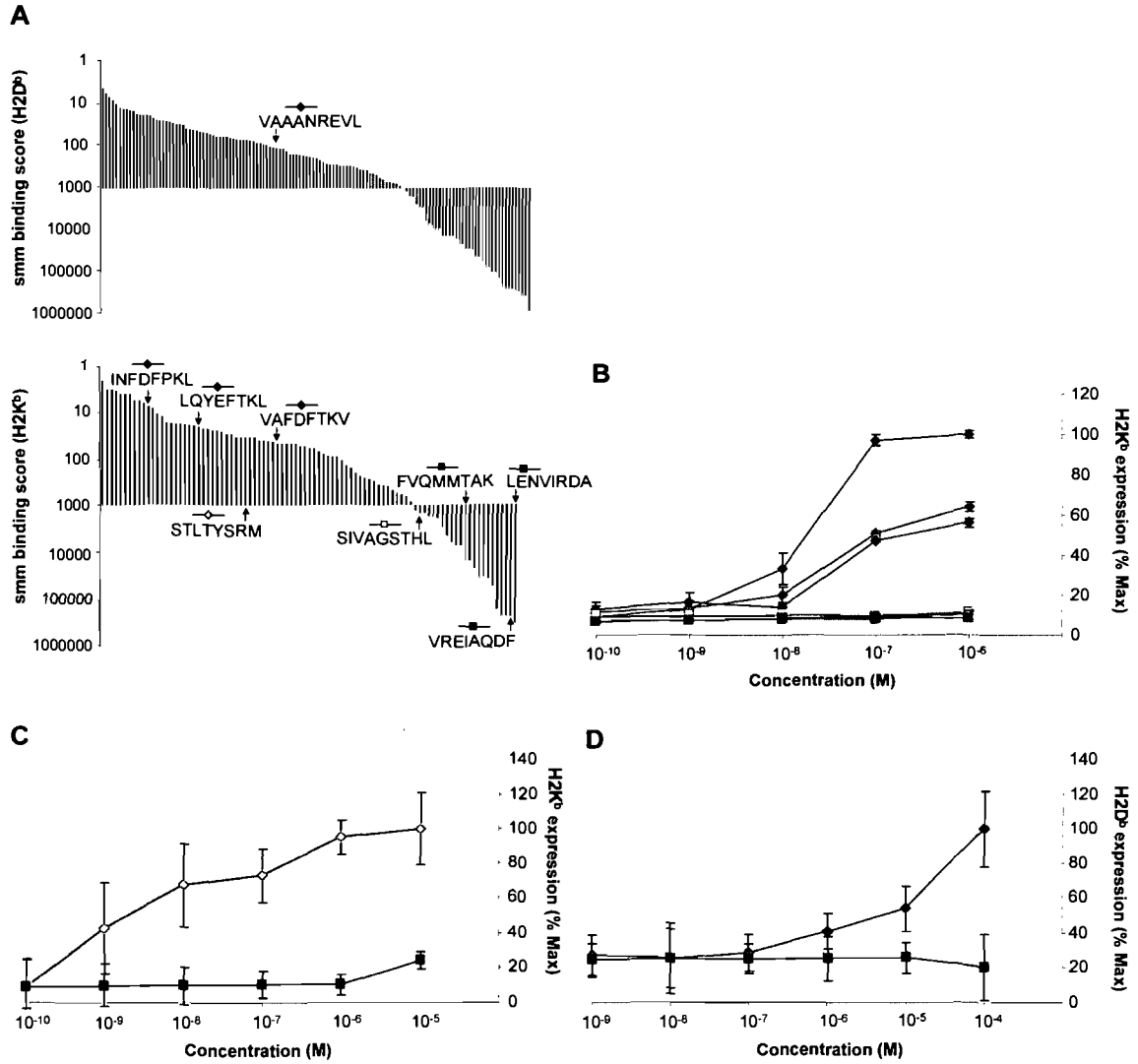

Figure 4.5: Evaluation of MHC binding affinity of 9 peptides. (A) For peptides eluted from WT EL4 cells, the y axis shows computed MHC binding scores determined with the smm computational method. The x axis cut at the selected binding thresholds. Each bar represents a sequenced peptide. Individual $H2K^b$- or $H2D^b$-associated peptides (red) and contaminant peptides (blue) were scored as illustrated. (B-D) We synthesized 9 peptides across the predicted affinity spectrum and tested their ability to bind $H2K^b$ or $H2D^b$. T2-K^b or T2-D^b cells maintained at 26°C for 18–20 h were pulsed with graded concentrations of the indicated peptides for 45 min at 26°C, washed, and incubated at 37°C for 4 h. Cells were stained with PE-conjugated anti-$H2K^b$ or FITC-conjugated anti-$H2D^b$ antibody and analyzed by flow cytometry. Mean ± SD of three experiments. Experimental results correlated perfectly with bioinformatic predictions: peptides classified as MHC I-associated did bind to MHC I while peptides classified as contaminants did not. (B) Stabilization of $H2K^b$ by seven peptides. (C,D) Stabilization of $H2K^b$ or $H2D^b$ by two peptides used in cytotoxicity assays (Figure 4.9). The RPQASGVYM peptide was used as negative control.

4.3.4 Global portrayal of the MIP repertoire of primary mouse thymocytes

In the next series of experiments, peptides eluted from primary mouse thymocytes by MAE were analyzed by on-line 2D-nanoLC-MS/MS. Of the sequenced peptides, 189 were classified as MHC I-associated based on the smm, SYFPEITHI and Rankpep thresholds selected above: 84 $H2D^b$-, 91 $H2K^b$-, 13 Qa2- and 1 Qa1-associated peptides (Table AII.III). Based on analyses described in the previous paragraph, we estimate that the false-positive rate for the 189 thymocyte peptides is < 2%. GO term enrichment analysis was performed on the 189 genes coding for those peptides by applying stringent criteria from 772 functional annotations (Figure 4.6A). In accordance with a study on HLA-B*1801-associated peptides [25], the MIP repertoire of primary thymocytes showed a modest but significant 2-fold enrichment in proteins located in the cytoplasm and the nucleus. More interestingly, we found an 11-to-19-fold enrichment in proteins related to cyclin and cyclin-dependent kinases (Ccng1, Cdkn1b, Chek1, Bccip, Cul2, Ccnd3, CcnF) which regulate the cell cycle. We also found an 11-to-17-fold enrichment in helicases (Dhx15, Ddx5, Ddx6, Hells, Ddx47), which are required for efficient and accurate replication, repair, and recombination of the genome.

4.3.5 MHC I-associated peptides derive preferentially from highly abundant mRNAs

We next sought to determine whether the MIP repertoire of thymocytes was molded at the mRNA level. To this end, we compared the relative abundance of two sets of thymic transcripts: (a) those encoding MHC I peptides eluted from primary thymocytes (Table AII.III) and (b) 36,182 transcripts encoded by the mouse genome [50]. We found a dramatic enrichment in highly abundant mRNAs among transcripts coding for MHC I peptides (Figure 4.6, B and C). Thus, although 9% of total mRNAs are expressed at high levels, 42% of those coding MHC I peptides did so (Figure 4.6C). Conversely, 62% of total mRNAs showed low abundance but only 20% of those coding MHC I peptides were expressed at low levels. Nonetheless, some MHC I peptides are coded by low abundance mRNAs (Figure 4.6, B and C). We hypothesized that low abundance mRNAs may contribute to the MIP repertoire because they code peptides with very high affinity for MHC I. Our data did not support that assertion. We found no correlation between the computed MHC I binding affinity of peptides and the abundance of their mRNAs (Figure

4.6D). MHC I peptides derived from low-abundance transcripts did not display superior computed MHC binding affinity. Although the ability to detect MHC I peptides is limited by the MS sensitivity (high attomole range) and dynamic range of detection (13 orders of magnitude on $\log_2$ scale of Figure 4.6), our results demonstrate that MHC I peptides derived preferentially, but not exclusively, from highly abundant mRNAs.

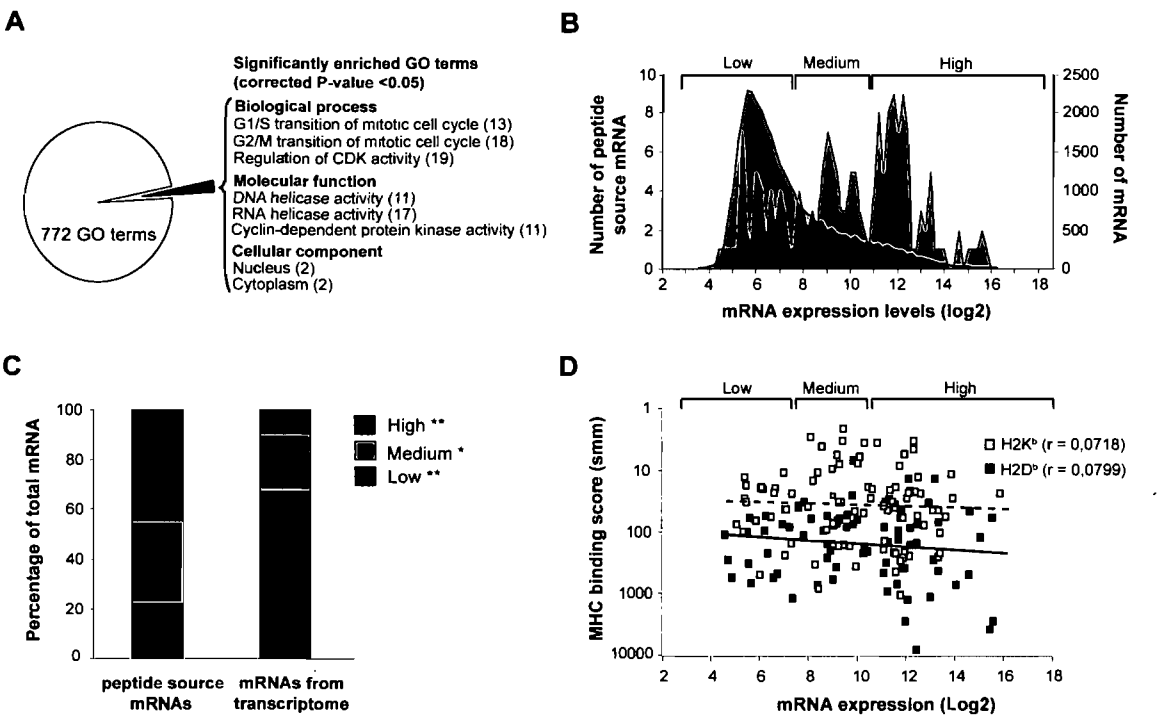

Figure 4.6: Analyses of genes and transcripts coding MHC I-associated peptides eluted from primary thymocytes. (A) GO term enrichment analysis of 189 genes coding MHC I peptides eluted from thymocytes. Exact P-values and global false-discovery rates were < 0.05 for each listed GO term. Values in parentheses indicate the fold enrichment relative to the whole mouse genome. (B) We compared the relative abundance of two sets of thymic transcripts using previously reported microarray data [50]: mRNAs coding MHC I peptides eluted from primary thymocytes (red), and 36,182 thymus-derived transcripts (gray). Original mRNA expression data on the x axis were plotted on a $\log_2$ scale. The y axes represent the number of transcripts for each sample set. (C) Frequency distributions for the two sets of thymic transcripts defined in B were plotted using a bin increment of 0.2. Three distinct mRNA expression groups are shown (low-, medium- and high-abundance mRNA, determined in an arbitrary manner). Graph shows the proportion of mRNAs with low- (black), medium- (red), and high- (blue) abundance among the two sets of transcripts. * $P < 0.05$, ** $P < 0.0001$. (D) Predicted MHC binding score (determined with smm) for peptides whose mRNA are expressed at low, medium or high level. Spearman linear correlation coefficient (r) was calculated for $H2K^b$- (dashed line; white squares) and $H2D^b$-associated (solid line; black squares) peptides.

4.3.6 Evidence that the MIP repertoire of thymocytes conceals a tissue-specific signature

Because MHC I peptides derived preferentially from highly abundant mRNAs, and abundance of discrete mRNAs varies among different tissues and organs, we hypothesized the MIP repertoire might harbor a tissue-specific signature. Using the SymAtlas database (http://symatlas.gnf.org/SymAtlas/) [50], we analyzed the relative expression in 58 mouse tissues of mRNAs encoding thymocyte MHC I-associated peptides (see Materials and methods for details on calculation). Remarkably, the mean expression of the 180 mRNAs coding thymocyte MHC I peptides was higher in the thymus than in all other tissues and organs (Figure 4.7A).

Large-scale quantitative analyses of transcriptional expression profiles across different tissues has revealed broadly expressed and tissue-specific groupings [50, 51]. To determine whether the MIP repertoire of thymocytes is imprinted by tissue-specific genes, we attributed a Z score to individual source mRNAs [52]. Genes with high Z score (from ~2.5 to 7) correspond to those that are preferentially overexpressed in the thymus. Among the 180 genes encoding thymocyte MHC I peptides, 30 showed a high Z score (Figure 4.7, B and C). The Z score distribution of genes encoding thymocyte MHC I peptides is illustrated in Figure 4.7B, and a heat map representation of the 30 genes with high Z scores is depicted in Figure 4.7C. Next, we estimated how additive removal of high Z score genes would affect the thymus specificity of the gene set. Thymus specificity of the gene set was lost after removal of the 30 high Z score genes (Figure 4.7D). In contrast, removal of up to 70 random genes did not affect the thymus specificity of the gene set. These results suggest that the MIP repertoire of thymocytes conceals a tissue-specific signature that is constituted by ~30 genes, i.e., ~17% of genes that are represented in the thymocytes' MIP repertoire.

Among the 30 genes with high thymic Z score, 16 are expressed, albeit to lower levels in other organs, whereas 14 genes are expressed almost exclusively in hematolymphoid organs: *Rhoh*, *Centb1*, *Cxcr4*, *Depdc1a*, *Foxp1*, *Dnmt1*, *9230105E10Rik*, *Cd3e*, *C330027C09Rik*, *Igtp*, *Mns1*, *Dock2*, *Actr2* and *Vps13d* (Figure 4.7C). In accordance with the concept that tissue-specific expression is indicative of gene function in mammals [51], we noted that hematolymphoid genes represented in the thymocytes' MIP repertoire

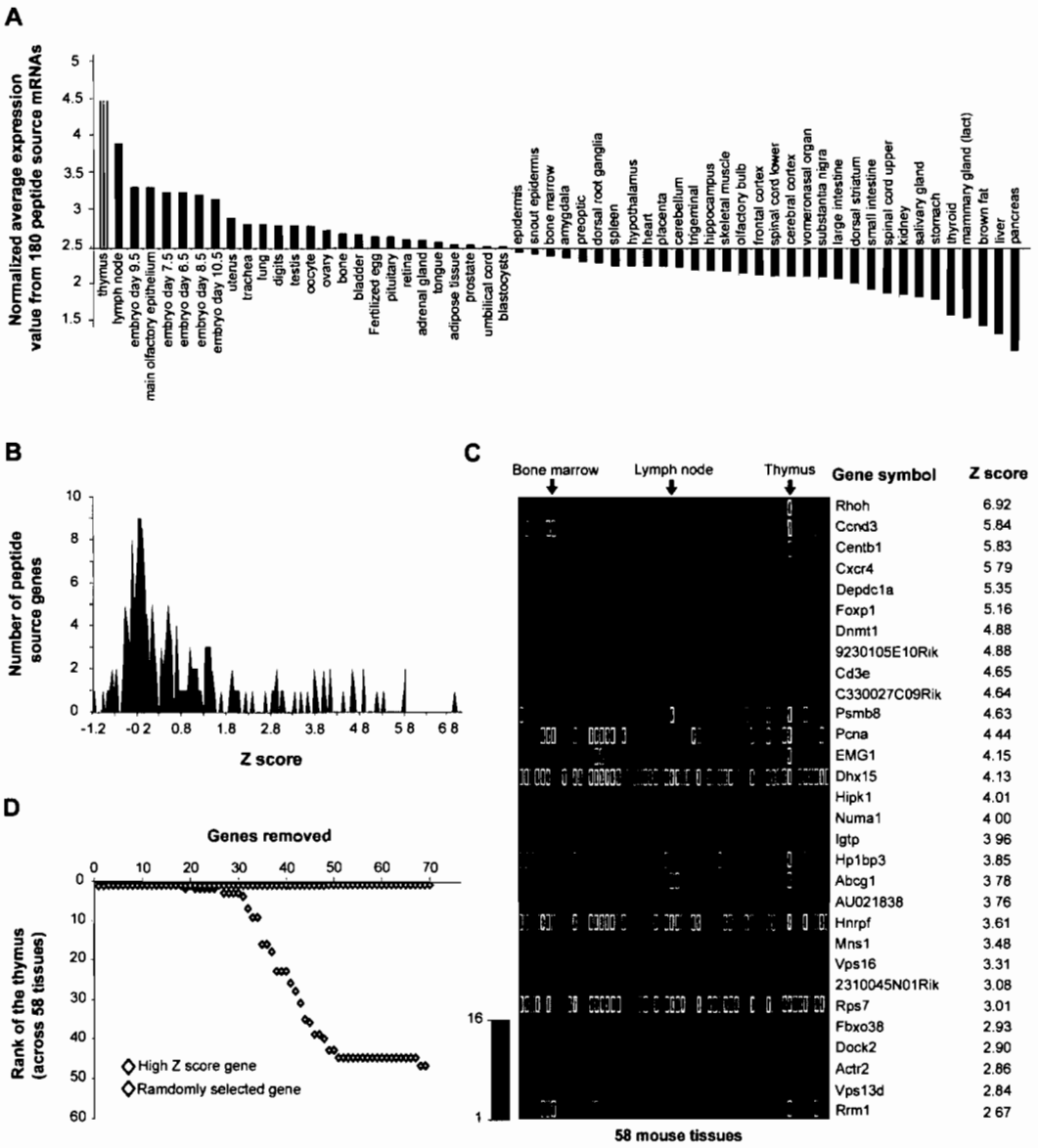

Figure 4.7: Peptide source mRNAs expression patterns reveal an organ-specific signature in the MIP repertoire of thymocytes. (A) Comparison of normalized mean expression values (y axis) across 58 different tissues (x axis) including the thymus (red). Normalized mean expression values were calculated as follows: average expression value from 180 peptide source genes/mean expression value of 36,182 transcripts for each particular tissue. Calculated values were ranked from left to right in a decreasing order. (B) Z scores were calculated for each of the 180 peptide source genes to identify transcripts preferentially expressed in the thymus. Graph shows frequency distributions of calculated Z scores with a bin increment of 0.05. (C) Heat map shows the relative mRNA expression in 58 tissues of the 30 peptide source genes with highest thymic Z scores. (D) High Z score genes determine the thymus specificity of the gene set encoding MHC I-associated peptides. Genes preferentially overexpressed in the thymus (high thymic Z scores; blue) were additively removed (x axis). Normalized mean expression values and thymus rank (y axis) were determined following removal of each individual gene. Removal of the 30 genes with a high thymic Z score had a drastic impact on thymus rank. Removal of up to 70 randomly selected genes (100,000 permutations; green) had no significant impact on thymus rank.

play critical roles in T-cell development. For example, *Rhoh* is important for positive thymocyte selection, *Cxcr4* for migration of T-cell progenitors, and *CD3e* for migration of T-cell development [53-55]. Further studies are needed to cogently assess the functional importance of genes that impart tissue specificity to the thymocytes' MIP repertoire. However, evidence suggests that many, if not all, of these genes are functionally important for thymocyte function. In conclusion, two major and related points can be made regarding the connection between the transcriptome and the MIP repertoire. First, the MIP repertoire is enriched in peptides derived from highly abundant transcripts. Second, our data suggest that thymocytes' MIP repertoire conceals a tissue-specific signature that derives from ~17% of MHC I-associated peptides.

4.3.7 The MIP repertoire of normal versus neoplastic thymocytes

Ultimately, the genesis of MHC I peptides must be regulated by mRNA translation and protein degradation by the proteasome [2, 49], two processes that are profoundly perturbed in neoplastic cells [56, 57]. To evaluate the impact of neoplastic transformation on the MIP repertoire, we compared the MIP repertoire of primary thymocytes from C57BL/6 female mice to that of neoplastic thymocytes. As a source of neoplastic thymocytes, we used *in vivo* grown EL4 cells. EL4 cells were originally derived from a C57BL/6 female mouse. Based on aforementioned experiments on the reproducibility of estimation of peptide abundance by MS analyses (Figure 4.2), peptides were considered to be differentially expressed when the fold difference in abundance was ≥ 2.5 ($P < 0.05$). Table AII.III and Table AII.IV present the complete list of peptides eluted from primary thymocytes and *in vivo* grown EL4 cells and their computed MHC binding score. Overall, 25% of MHC I peptides were differentially expressed on normal versus neoplastic thymocytes (Table 4.2 and Figure 4.8A). Thus, 22 peptides were underexpressed and 21 were overexpressed on neoplastic relative to normal thymocytes (Table 4.2). Differentially expressed peptides derived from genes implicated in several biological processes, such as cell cycle progression, apoptosis, signal transduction, cytoskeleton assembly, differentiation, as well as regulation of transcription and translation (Table 4.2). As an example, the X-linked lymphocyte-regulated 3 (*Xlr3a/b*) gene encoded a peptide found only on EL4 cells (Figure 4.8B and Table 4.2). *Xlr* genes are important for T cell differentiation and are overexpressed in several lymphoid malignancies [58]. In contrast, an MHC I peptide from cyclin-dependent kinase inhibitor 1B (*Cdkn1b*) was underexpressed on neoplastic cells

Table 4.2: Peptides differentially expressed on EL4 cells versus primary thymocytes.

Functional Classification	Gene Symbol	Gene ID	Function	Sequence	P-value	EL4/ Thy
Transcription	*Pa2g4*[a]	18813	Growth regulation	AQFKFTVL	0.02	4.6
	Top2a[a]	21973	DNA topoisomerase	NSMVLFDHV	0.02	15.6
	Dnmt1[a]	13433	DNA methylation	LSLENGTHTL	0.02	8.9
	Pfdn5[a]	56612	Protein binding/Folding	SMYVPGKL	0.04	16.6
	Per1[a]	18626	Transcription regulator	YTLRNQDTF	0.04	5.0
	Foxp1[a]	108655	Transcription factor	QQLQQQHLL	0.01	-4.3
	Bach2[a]	12014	Transcription factor	EQLEFIHDI	0.02	-20.2
	Pbrm1	66923	DNA binding	SQVYNDAHI	0.02	-3.5
	Ddx5[a]	13207	RNA helicase	NQAINPKLLQL	0.02	-5.7
	Jhdm1d	338523	Histone demethylase	SSIQNGKYTL	0.02	-4.9
Cell Differentiation	*Ptpn6*[a]	15170	Key role in hematopoiesis	AQYKFIYV	0.03	3.3
	Mark2	13728	Maintenance of cell polarity	ASIQNGKDSL	0.03	2.7
	Pi3kap1[a]	83490	Role in BCR-mediated Pi3K activation	YGLKNLTAL	0.001	-21.8

Table 4.2: Peptides differentially expressed on EL4 cells versus primary thymocytes. **(continued...)**

Functional Classification	Gene Symbol	Gene ID	Function	Sequence	P-value	EL4/ Thy
	Xlr3a/b	22446	Role in lymphocyte development	VAAANREVL	0.02	84.9
	RhoH[a]	74734	Small GTPase/thymocyte maturation	YSVANHNSFL	0.02	-3.1
Cell Cycle	*Rcc2*	108911	Required in mitosis and cytokinesis	AAYRNLGQNL	0.04	27.0
	Cdkn1b[a]	12576	Involved in G1 arrest	FGPVNHEEL	0.003	-3.1
Apoptosis	*Sgk*[a]	20393	Response to DNA damage stimulus	STLTYSRM	0.02	9.7
	Pdcd10[a]	56426	Role in apoptotic pathways	ILQTFKTVA	0.001	-4.3
Signal Transduction	*Cxcr4*[a]	12767	Receptor for the chemokine CXCL12/SDF1	VVFQFQHI	0.003	2.6
	CD97[a]	26364	Involved in adhesion and signaling process	KLLSNINSVF	0.03	-4.6
Translation	*Eif3s10*[a]	13669	Translation initiation factor	QSIEFSRL	0.02	3.2
	Eif3s2[a]	54709	Translation initiation factor	FGPINSVAF	0.04	5.3
Transport	*Tmed9*	67511	Intracellular protein transport	VIGNYRTQL	0.02	-6.9
	Copb1	70349	ER to Golgi vesicle-mediated transport	IALRYVAL	0.01	2.9
Cytoskeleton	*Tmod1*	21916	Cytoskeleton organization	SSIVNKEGL	0.004	10.5

Table 4.2: Peptides differentially expressed on EL4 cells versus primary thymocytes. **(continued...)**

Functional Classification	Gene Symbol	Gene ID	Function	Sequence	P-value	EL4/ Thy
	Mylc2b	67938	Cytoskeleton organization and biogenesis	SLGKNPTDAYL	0.01	4.2
	Mns1	17427	Involved in cell division and motility	KIIEFANI	0.01	-12.4
	Krt5	110308	Strutural constituent of cytoskeleton	AAYMNKVEL	0.04	-106.7
	Krt7	110310	Strutural constituent of cytoskeleton	AAYTNKVEL	0.003	-121.5
Miscellaneous	*Igtp*[a]	16145	IFN gamma induced GTPase	IVAENTKTSL	0.001	-15.5
	Nup205	70699	Outer membrane exporter porin	VNNEFEKL	0.005	-4.6
	Dhx15[a]	13204	RNA helicase	TLLNVYHAF	0.02	-11.5
	Stk11ip	71728	Serine/Threonine kinase	SALRFLNL	0.01	-3.9
	Pde2a	207728	Catalytic activity	IKNENQEVI	0.02	16.0
	Gtpbp4[a]	69237	GTP binding	QILSDFPKL	0.03	2.6
	Narfl	67563	Prelamin recognition	VAYGFRNI	0.01	8,7
	Rmnd5a	68477	Hypothetic role for meiotic nuclear division	WAVSNREML	0.02	-6.2
Unknown	*Specc1*	432572		TSLAFESRL	0.01	7.1

Table 4.2: Peptides differentially expressed on EL4 cells versus primary thymocytes. **(continued...)**

Functional Classification	Gene Symbol	Gene ID	Function	Sequence	P-value	EL4/ Thy
	Ccdc41	77048		AQVENVQRI	0.01	-2.8
	2900073G15Rik	67268		SMGKNPTDEYL	0.05	3.8
	D14Ertd436e	218978		SQHVNLDQL	0.04	-2.7
	9230105E10Rik	319236		FISDVEHQL	0.01	-23.2

Gene ID and Gene Symbol description refer to NCBI gene entries. Fold change and P-values were calculated from biological replicate experiments (n = 3): Functional classification is based upon bibliographic searches. [a]Indicate genes that are involved in neoplastic transformation (See Annexe 1, Table AII.VI for references).

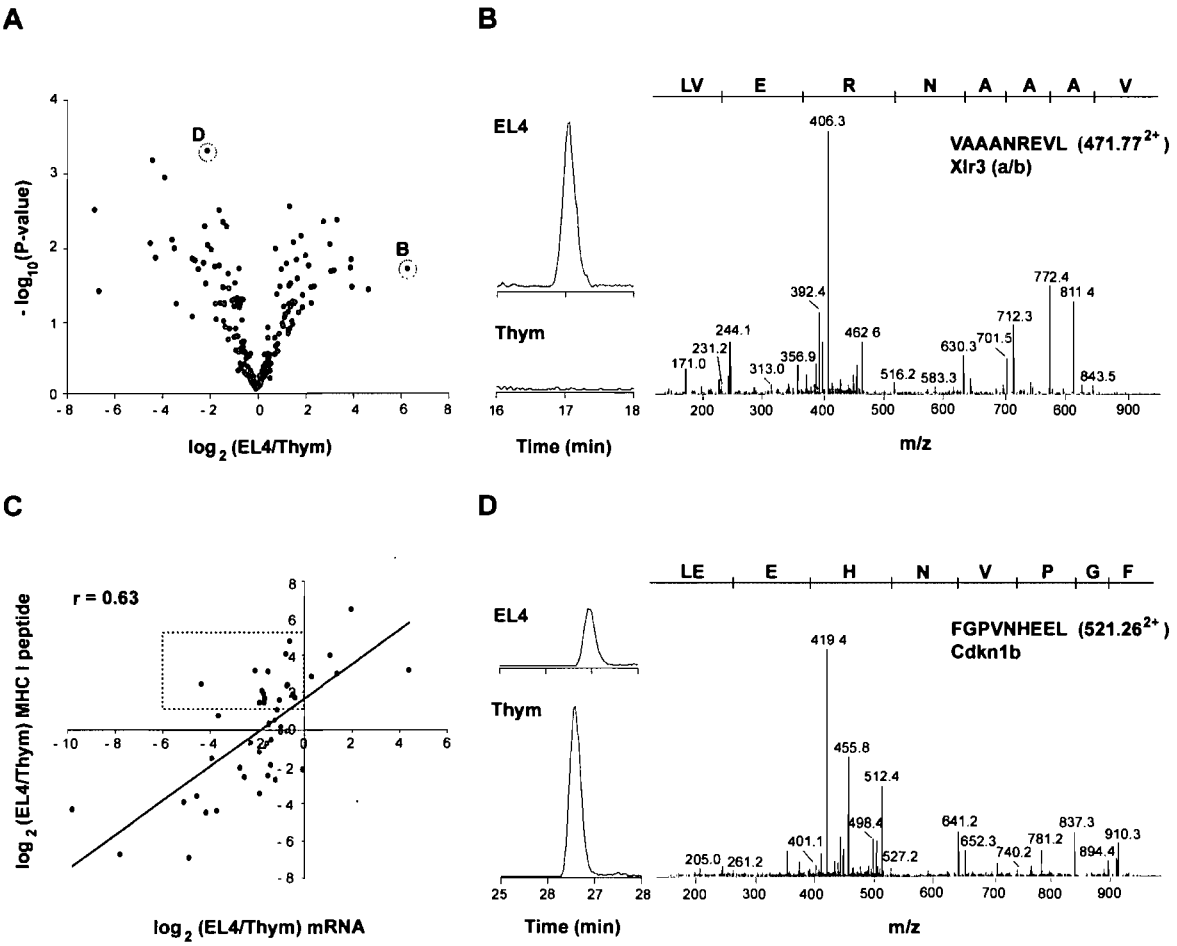

Figure 4.8: Relative quantification of differentially expressed MHC I peptides and source mRNAs from thymocytes and EL4 cells. (A) Volcano Plot representation illustrates MHC I peptides reproducibly detected across biological replicates (n = 3). Peptides over- and underexpressed on EL4 cells relative to thymocytes ($P \leq 0.05$; fold change ≥ 2.5) were highlighted in blue and red, respectively. MS/MS spectra of circled peptides are shown in B and D. (B and D) Illustration of two differentially expressed MHC I peptides. Reconstructed ion chromatograms show differential abundance for m/z 471.77^{2+} (VAAANREVL) and 521.26^{2+} (FGPVNHEEL) in EL4 cells versus thymocytes. MS/MS spectra confirm MHC I peptide sequences and the identification of the cognate source proteins. (C) Scatter plot shows the correlation between relative expression of mRNA and that of MHC I peptide. Expression ratios for source mRNA (x axis) and MHC I peptide (y axis) between EL4 cells and thymocytes were plotted on a $\log_2$ scale for 47 pairs. A Spearman correlation coefficient was calculated from the linear regression. MHC I peptides overexpressed in EL4 cells or normal thymocytes are highlighted in blue and red, respectively; peptides that were not differentially expressed are shown in gray. Dashed box includes peptides whose overexpression on EL4 cells did not correlate with increased mRNA levels of their source protein.

(Figure 4.8D and Table 4.2). *Cdkn1b* is known to act as a potent tumor suppressor gene in a variety of cancers [57]. Remarkably, ~50% of differentially expressed peptides derived from genes that are known to be involved in tumorigenesis (Table 4.2 and Table AII.VI). For exemple, ten differentially expressed peptides (*Bach2*, *Cdkn1b*, *Cxcr4*, *Eif3s2*, *Eif3s10*, *Igtp*, *Pa2g4*, *Pi3kap1*, *Ptpn6* and *Sgk*) originated from genes related to the PI3K-AKT-mTOR pathway, which is the oncogenic signaling pathway most commonly targeted by genomic aberrations in cancer [59, 60].

4.3.8 Genesis of peptides overexpressed on tumor cells

An important question is whether differential expression of MHC I peptides on neoplastic relative to normal thymocytes correlates with changes in mRNA levels of source transcripts. To test this, we selected 19 peptides overexpressed on EL4 cells relative to primary thymocytes, 15 that were underexpressed, and 13 that were not differentially expressed. Then, we analyzed the level of expression of their source mRNAs in neoplastic versus primary thymocytes using quantitative real-time PCR. Scatterplot representation of the correlation between relative mRNA expression and relative MHC I peptide expression is depicted in Figure 4.8C. From the linear regression, a Spearman coefficient of 0.63 was calculated, showing a significant but moderate correlation between peptide and mRNA expression ratios. The strength of the correlation was conspicuously decreased by a set of 14 peptides that were more abundant on neoplastic cells but whose mRNA levels were not overexpressed (Figure 4.8C, dotted box). Exclusion of these 14 genes increased the correlation coefficient to 0.78. Remarkably, for the 19 peptides overexpressed on EL4 cells, increased transcript levels were present in EL4 cells in only 5 cases (Figure 4.8C). Peptides overexpressed on neoplastic cells are particularly important because they can be used as targets in cancer immunotherapy [61, 62]. Thus, the salient finding here is that 74% (14 of 19) of peptides overexpressed on EL4 cells would have been missed by studies of mRNA expression levels. An important implication is that, at least in our model, overexpression of MHC I-associated peptides on neoplastic cells generally entails posttranscriptional mechanisms.

4.3.9 Testing the immunogenicity of peptides overexpressed on neoplastic cells

Finally, we wished to determine whether peptides overexpressed on neoplastic thymocytes (*in vivo* grown EL4 cells) would be able to elicit specific CD8 T cell response. To test this, we used two peptides: (a) STLTYSRM, which is derived from serum/glucocorticoid regulated kinase (Sgk) and presented by $H2K^b$, and (b) VAAANREVL derived from X-linked lymphocyte-regulated 3 (Xlr3a/b) and presented by $H2D^b$. One important difference between the two peptides is that VAAANREVL was not found on primary thymocytes (fold difference relative to EL4 cells ≥ 85), whereas low levels of STLTYSRM were present on primary thymocytes (fold difference = 10) (Table 4.2 and Table AII.V). TAP-deficient T2-D^b and T2-K^b cells were first incubated with titrated amounts of synthetic peptides to evaluate their ability to bind and stabilize $H2K^b$ and $H2D^b$. In contrast to the L^d-restricted peptide RPQASGVYM, both STLTYSRM and VAAANREVL peptide loading resulted in an increase in cell surface expression of H2-K^b or H2-D^b (Figure 4.5, C and D), thereby confirming their ability to bind their respective MHC allele.

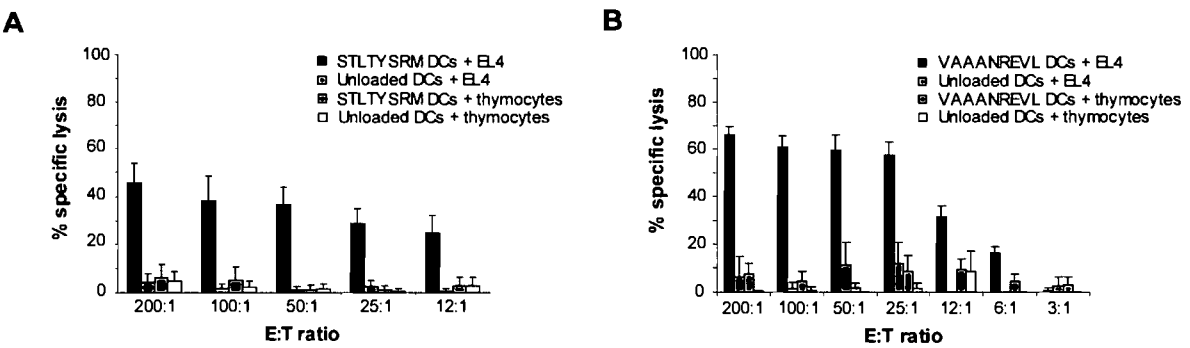

Figure 4.9: Splenocytes primed against peptides overexpressed on EL4 cells selectively kill EL4 cells. Mice were immunized with DCs coated with (A) STLTYSRM or (B) VAAANREVL peptide. Splenocytes from primed mice were restimulated *in vitro* with the corresponding peptide for 6 days, and tested for *in vitro* cytotoxic activity against CFSE-labeled target cells (EL4 cells and primary mouse thymocytes) at different E/T ratios. Number of effectors represents the number of unfractionated splenocytes used in the cytotoxicity assay. Mice immunized with unloaded DCs were used as negative control. Data represent the mean ± SD for four mice per group.

We next immunized C57BL/6 mice with DCs coated or not with VAAANREVL and STLTYSRM synthetic peptides. Splenocytes from immunized mice were tested for *in vitro*

cytotoxicity against primary thymocytes and EL4 target cells that were not loaded with exogenous peptides. Splenocytes from mice primed with unloaded DCs showed no cytotoxic activity. In contrast, splenocytes primed with coated DCs killed EL4 cells, but not primary thymocytes (Figure 4.9). Thus, peptides overexpressed by 10 to $\geq$ 85 fold on neoplastic cells elicited specific cytotoxic activity against endogenously presented epitopes.

4.4 Discussion

We have developed a novel method for high-throughput analysis of MHC I peptides and performed a comprehensive study of the MIP repertoire of normal and neoplastic thymocytes. Our studies yielded important insights into the genesis of the MIP repertoire and how it is modified by neoplastic transformation.

4.4.1 A novel method for high-throughput, MS-based analysis of MHC I peptides

In the last few years, high-throughput screening methods coupled with bioinformatics tools have helped to figure out the complexity of biological matrices. Thus, emerging technologies in mass spectrometry have led to the development of peptide detection algorithms that can be used, in combination with segmentation analyses, to compare unlabeled peptide populations (unpublished data) [36]. This powerful approach allowed us to obtain an accurate quantification of native unlabeled MHC I peptides without any chemical or metabolic labeling modifications. Because preparation of samples does not require different purification steps, higher sensitivity can be achieved from limiting the amount of materials compared to chemical derivatization where peptide recovery can be affected by variable modification yields or side reaction products [31]. Thus, by combining our MS strategy with MAE, we were able to generate a comprehensive portrayal of the MIP repertoire of EL4 cells from $<10^8$ cells. The ability to identify large numbers of MHC I peptides from limiting amounts of cells is a noteworthy advantage. Moreover, our quantification method can be used to analyze any type of cell population, whereas metabolic labeling strategies can only be applied to cell culture model systems. Nevertheless, low abundance peptides (<100 copies/cell) remain challenging to identify in view of the sensitivity of present high-throughput, MS-based methods. The portrayal of the MIP repertoire described in this study and others is far from complete, and low abundance peptides that may be biologically relevant could remain elusive to current detection methods.

We took advantage of EL4 variant cell lines (WT, β_2m^- mutant and β_2m^+ transfectant) to unambiguously discriminate between MHC I-associated peptides and

contaminants in peptide mixtures obtained by MAE. A large proportion of contaminants was made of long peptides derived from the C-terminal end of source proteins (unpublished data). We noted that some of these contaminants have previously been considered as MHC I-associated peptides in studies where peptides were obtained by MAE or immunoaffinity purification [23, 63]. The need to have an MHC I-deficient negative control to distinguish MHC I peptides from contaminants would be a cumbersome limitation for analysis of primary cells. However, we showed that the use of computational models such as SYFPEITHI and smm obviated this need. Thus, thresholds used herein yielded false-positive and –negative rates of ~ 2% in identification of MHC I peptides.

4.4.2 The MIP repertoire of primary cells

High-throughput analysis of peptides obtained by MAE provided us with a global portrayal of peptides presented by different MHC I allelic products (in this study: $H2D^b$, $H2K^b$, Qa1 and Qa2). We found that, with rare exceptions, discrete MHC I molecules presented peptides derived from different sets of source proteins (Table AII.II and AII.III). A corollary is that expression of multiple MHC I allelic products (a consequence of gene duplication and diversification [64]) favors representation of largely nonoverlapping sets of source proteins in the MIP repertoire. By integrating global profiling of the mouse protein-encoding transcriptome [50] with the MIP repertoire of thymocytes, we found that the thymocytes' MIP repertoire is enriched in peptides derived from highly abundant transcripts. Furthermore, our data suggest that the repertoire of MHC I-associated peptides conceals a tissue-specific signature that derives from ~17% of genes represented in the MIP repertoire. Cogent evaluation of this exciting concept will require comprehensive analyses of MHC I-associated peptides eluted from various tissues and organs. Why would the MIP repertoire show a stronger correlation with the transcriptome (this study) than the proteome [28]? Probably because the proteome is enriched in slowly degraded proteins (with an average $t_{1/2}$ of >1,000 min) whereas the MIP repertoire originates mainly from rapidly degraded proteins (with an average $t_{1/2}$ of ~10 min) that have recently been translated and degraded [28, 49]. In other words, MHC I molecules sample what is being translated rather than what has been translated [49]. Considering the pervasive roles of MIPs, particularly in CD8 T cell development and function, it will be extremely interesting to determine whether the MIP repertoire of specialized cell types does conceal a tissue-specific signature. Of

particular interest are thymus cortical epithelial cells, which support thymocyte positive selection, and thymus medullary epithelial cells, which express promiscuous transcripts involved in tolerance induction. Besides, although the MIP repertoire is enriched in peptides derived from highly abundant mRNAs, some peptides presented by MHC class I molecules derive from low abundance mRNAs (Figure 4.6). How can low-abundance transcripts successfully compete with more abundant transcripts for representation in the MIP repertoire? Attractive possibilities would be that successful transcripts generate more DRiPs than others or that they are translated by special "immunoribosomes" [33, 49].

The proteasome, the primary source of MHC I peptides, is much more ancient than the MHC. The MHC appeared in gnathostomes while proteasomes are found in all eukaryotes. Fundamental functions of the proteasome are to regulate cell cycle, proliferation and apoptosis [65]. In line with this, we found that the MIP repertoire of primary thymocytes was enriched in peptides derived from cyclins, cyclin-dependent kinases and helicases (Figure 4.6A). These highly conserved proteins regulate cell proliferation, cell cycle arrest and apoptosis. Peptides derived from cyclin, cyclin-dependent kinase and helicase gene families have also been found in other studies of MHC I peptides [9, 25, 63]. We surmise that the presence of these peptide families in the MIP repertoire may represent an imprint of primordial functions of the proteasome. Nevertheless, an alternative hypothesis would be that overrepresentation of peptides-derived proteins regulating proliferation and apoptosis is a thymocyte-specific feature because thymocytes display particularly high rates of proliferation and apoptosis.

4.4.3 The MIP repertoire of neoplastic cells

Neoplastic transformation is associated with many genomic and proteomic changes. How these changes may impinge on the MIP repertoire and thereby be perceived by CD8 T cells, is a fundamental question in cancer immunology that can be addressed directly only by MS-based analysis of the MIP repertoire. Weinzierl *et al.* recently reported a seminal study in which they integrated mass spectrometry and microarray data on renal cell carcinomas and autologous normal kidney tissues [9]. They found a poor correlation between changes in peptide expression and changes in the abundance of mRNA levels in normal versus cancer cells ($r = 0.32$). We found a better correlation between mRNA levels

and expression of corresponding MHC I peptides in normal and neoplastic thymocytes (Figure 4.8C; $r = 0.63$). We surmise that the stronger correlation in our case may be caused by estimation of transcript levels with quantitative real-time PCR analyses rather than microarrays. Indeed, quantitative real-time PCR provides a more accurate estimation of quantitative differences than microarrays [66]. However, our data support the main conclusion of Weinzierl *et al.*, which is that a majority of peptides overexpressed on neoplastic cells (74% in our case) would have been missed by estimation of transcript levels. Together with data from Weinzierl *et al.*, our results suggest that in general, overexpression of MHC I peptides on neoplastic cells is caused by posttranscriptional mechanisms and can be detected only by MS-based expression profiling approaches. Further studies on diverse cancer types are needed to test the generality of these observations. Relevant mechanisms may include dysregulation of microRNA expression, protein translation, and proteasomal degradation [67-69].

Our data suggest that the MIP repertoire gives a unique perspective into the mechanisms of carcinogenesis. Approximately 50% of peptides differentially expressed on normal versus neoplastic cells are coded by genes involved in neoplastic transformation (Table 4.2 and Table AI.VI). It is quite remarkable that ten of the differentially expressed peptides derived from genes related to the PI3K-AKT-mTOR pathway. The PI3K-AKT-mTOR pathway is the most prominent pathway regulating protein translation, cell growth, and cell proliferation, and is activated in most cancers [59]. In addition, our data lead us to speculate that the MIP repertoire might give unique insights into acquired epigenetic abnormalities that are so important in cancer. Indeed, *Dnmt1*, which regulates DNA methylation and prevents genomic instability [70], is expressed at high levels in thymocytes (Figure 4.7C). We found that a *Dnmt1*-derived peptide was overexpressed ~9 fold on neoplastic relative to primary thymocytes (Table 4.2). This raises the enticing possibility that enhanced proteasomal degradation of Dnmt1 might be responsible for the genomic instability of EL4 cells. Finally, we were able to generate specific cytotoxic T cells responses against EL4 cells by priming WT mice with two peptides overexpressed on EL4 cells (Figure 4.9). The finding that efficient cytotoxic responses could be elicited against endogenously expressed tumor peptides identified with our MS-based expression analyses validates the analytical potential of the present approach. Though further work is needed to evaluate the therapeutic value of vaccination with these peptides, our results strongly

support the concept that high-throughput analysis of the MIP repertoire is a promising discovery platform for the identification of immunogenic tumor-associated peptides [27, 61, 62].

4.5 Materials and methods

4.5.1 Chemicals and materials

Citric acid, aprotinin, iodoacetamide and sodium phosphate dibasic (Na_2HPO_4) were purchased from Sigma-Aldrich; high performance liquid chromatography-grade water, methanol (MeOH) and acetonitrile (ACN) were obtained from Fisher Scientific; formic acid (FA) and ammonium acetate were purchased from EM Science; fused-silica capillaries were obtained from Polymicro Technologies; and teflon and PEEK tubing were purchased from Supelco. The Jupiter Proteo C12 4 μm material used for packing homemade precolumn and column was obtained from Phenomenex. Strong cation exchange (SCX) material was purchased from PolyLC.

4.5.2 Cell lines and flow cytometry

The EL4 (β_2m^+ wildtype), C4.4-25$^-$ (β_2m^- mutant) and E50.16$^+$ (β_2m^+ transfectant of C4.4-25$^-$) cell lines were provided by Dr. R. Glas (Karolinska Institutet, Karolinska University Hospital, Huddinge, Sweden). T2-K^b and T2-D^b cell lines (gifts from S. Joyce, Vanderbilt University School of Medicine, Nashville, TN) were maintained as previously described [71]. MHC class I molecules at the cell surface were stained with PE-conjugated anti-H-2K^b (clone AF6-88.5, BD Pharmingen), FITC-conjugated anti-H-2D^b (clone KH95, BD Pharmingen), and biotin-conjugated anti-Qa-2 (clone 1-1-2, BD Pharmingen) and analyzed on a BD LSR II flow cytometer using FACSDiva software (BD Biosciences).

4.5.3 Peptide extraction and mass spectrometry analysis

Freshly isolated thymocytes were obtained from 4-to-6-wk-old C57BL/6 female mice purchased from The Jackson Laboratory. All mice were maintained under specific pathogen-free conditions according to the standards of the Canadian Council on Animal Care. Experimental protocols were approved by the Comité de Déontologie Animale of the Université de Montréal. Thymocytes were separated from stroma according to standard procedures [72]. For isolation of EL4 cells grown *in vivo*, mice were injected

intraperitoneally with 200,000 EL4 cells, and ascites fluid was harvested 17 days later. Purity of EL4 cells was assessed by flow cytometry [73]. Three biological replicates were prepared and analyzed for normal thymocytes and EL4 cells. Cell surface MHC I peptides were isolated from viable cells as previously described, using a slightly modified protocol [37]. In brief, 4 mL of citrate-phosphate buffer at pH 3.3 (0.131 M citric acid/0.066 M Na_2HPO_4, NaCl 150 mM) containing aprotinin and iodoacetamide (1:100) was added to each flask, and cell pellets were resuspended by gentle pipetting for 1 min to denature MHC I complexes. Cell suspensions were then pelletted and the resulting supernatant was isolated. Peptides extracts were desalted using Oasis HLB cartridges (30mg; Waters) and bound material was eluted with 1 mL H_2O/80% MeOH/0.2% FA (vol/vol) and diluted to H_2O/40% MeOH/0.2% FA (vol/vol). Peptides were then passed through ultrafiltration devices (Amicon Ultra; Millipore) to isolate peptides < 5,000 Da and to remove β_2m proteins. The resulting flowthrough was then lyophilized and stored at -30°C or -80°C until analysis.

To extract equivalent amount of MHC I peptide from EL4 cells and normal thymocytes, we assessed the amount of MHC I molecules on both cell populations using previously described methods [74]. Equivalent amount of MHC I material from neoplastic (8×10^7) and normal (2.3×10^9) thymocytes were separated by on-line 2D separation (SCX/C12 Jupiter Proteo 4 µm) using an Eksigent nanoLC-2D system. Samples were diluted in H_2O/2% ACN/0.2% FA before LC-MS analyses. The homemade SCX column (0.3 mm i.d. X 45 mm) was connected directly to the switching valve. During sample loading, SCX column was positioned off-line of C_{12} precolumn to remove interfering species. Salt fractions (10 µL each) were loaded on SCX column at 5 µL/min for 6 min to sequentially elute peptides onto the C_{12} precolumn using pulsed fractions of 0, 75, 300, 1000 mM of ammonium acetate (pH 3.0). A 69-min gradient from 3-60% acetonitrile (0.2% FA) was used to elute peptides from homemade reversed-phase column (150 µm i.d. X 100 mm) with a flow rate set at 600 nL/min. On-line 2D-nanoLC-MS system was used to provide enhanced selectivity, as well as a higher capacity to detect low abundance MHC I peptides. To achieve high mass accuracy and sensitivity from a limited amount of starting material, abundance profiles and tandem MS experiments were performed simultaneously using a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a

nanoelectrospray ion source (voltage 1.5-1.7 kV). MS scans were acquired in the FT mode (Orbitrap) with a resolution set at 60,000 (m/z 400). Each full MS spectrum was followed by three MS/MS spectra (four scan events), where the three most abundant multiply charged ions were selected for MS/MS sequencing. Tandem MS experiments were performed using collision-induced dissociation (CID) in the linear ion trap (IT mode). Target ions already selected for MS/MS fragmentation were dynamically excluded for 80 seconds, and a minimal intensity value of 10,000 was fixed for precursor ion selection.

4.5.4 Peptide detection and clustering

Raw data files generated from the Orbitrap were processed using in-house peptide detection software (Mass Sense) to identify all ions according to their corresponding m/z values, charge state, retention time and intensity. From this process, lists of detected peptides ions were generated to define individual LC-MS analysis. A user-defined intensity threshold above the background noise (typically between 5,000–15,000) was fixed to limit false-positive identification. Segmentation analyses were performed across sample sets using hierarchical clustering to generate lists of nonredundant peptide clusters [35]. User-specified tolerances were fixed to typically ± 0.04 m/z, ± 1.5 min and ± 1 fraction for 2D-LC-MS Orbitrap experiments. Identification files from Mascot were converted in excel format (Microsoft) and sequenced peptides were aligned with their corresponding peptide clusters using in-house clustering software.

4.5.5 MS/MS sequencing and protein identification

The data were searched against International Protein Index (IPI) mouse database using Mascot (Matrix Science) search engine. The mass tolerance on precursor and fragment ions were set to ± 0.1 Da and ± 0.4 Da, respectively. Searches were performed without enzyme specificity. All relevant hits were inspected manually for mass accuracy and to validate sequence assignment according to the observation of consistent b-, and y-types fragment ion series. From MHC I peptide structures identified, 85% were sequenced at least twice across replicate injections, and all had mass accuracy within 30 ppm of the theoretical values (without the use of a lock mass). To estimate the false-positive rate

across non-MHC I-associated peptide identification, β_2m-deficient EL4 samples were searched against a concatenated forward/reverse IPI mouse database to establish a cutoff score threshold (>45) for a false-positive rate of < 2%. All identified MHC I peptides were blasted against nonredundant NCBInr (restricted to mouse entries) to identify their corresponding source proteins (Gene ID).

4.5.6 Quantitative real-time PCR

Freshly isolated thymocytes and EL4 cells were homogenized in TRIzol RNA preparation reagent (Invitrogen), and total RNA was isolated as instructed by the manufacturer. Then 1 μg of total RNA was converted to cDNA using random hexamer priming (Thermoscript RT-PCR System; Invitrogen). Gene expression level was determined using primer and probe sets from Universal ProbeLibrary (Table AII.VII) https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp. PCR reactions for 384-well plate formats were performed using 2 μl of cDNA samples (50 ng), 5 μl of the TaqMan PCR Master Mix (Applied Biosystems), 2 μM of each primer and 1 μM of the Universal TaqMan probe in a total volume of 10 μl. The ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems) was used to detect the amplification level and was programmed to an initial step of 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. All reactions were run in triplicate, and the average values were used for quantification. The mouse GAPDH or 18S ribosomal RNA were used as endogenous controls.

4.5.7 Microarray dataset cross comparison

Normalized mRNA expression data [50] (http://symatlas.gnf.org/SymAtlas/) used in this study were visualized with TIGR MultiExperiment Viewer (http://www.tigr.org/software/microarray.shtml). Normalized average expression values were calculated as follows for each tissue: average expression value of 180 peptide source genes / average expression value of 36,182 transcripts.

4.5.8 Expression analysis systematic explorer analysis and statistical methods

The Expression Analysis Systemic Explorer (EASE) software [75] was downloaded from the Database for Annotation, Visualization, and Integrated Discovery (http://david.niaid.nih.gov/david/ease.htm) [76]. For gene enrichment analysis, statistical P-values and corrected P-values were calculated from the Fisher exact test and the global false-discovery rate, respectively.

4.5.9 Z score transformation

Raw intensity data for each tissue were $\log_2$ transformed and used for the calculation of Z scores. Z scores were calculated by subtracting the overall average tissue intensity (for a single gene) from the raw intensity data within a given tissue, and dividing that result by the SD of the overall tissue intensities: Z score = (intensity G_{Th} – mean intensity $G_{T1...Tn}$)/ SD $G_{T1...Tn}$ where G_{Th} is any gene on the microarray from the thymus tissue and T1...T*n* represents the aggregate measure of all tissues.

4.5.10 Peptide-binding and *in vitro* cytolytic assay

Peptides were synthesized by the Centre Hospitalier de l'Université Laval Research Center (Quebec) and purified by high performance liquid chromatography (purity > 90%). Rescue of class I expression in T2-D^b and T2-K^b cells by allele-specific peptides was determined as previously described [77]. Bone-marrow derived DCs were generated as previously described [78]. On day 9 of culture, peptides were added at a concentration of 2 x 10^{-6} M and incubated with DCs for 3h at 37°C. For mice immunization, 10^6 peptide-pulsed DCs were injected i.v. in C57BL/6 females at day 0 and 7. At day 14, splenocytes were harvested from the spleens of immunized mice and depleted of red blood cells using 0.83% NH_4CL. Cells were plated at 5 x 10^6 cells/well in 24-well plates and restimulated with 2 x 10^{-6} M peptide at 37°C. After 6 days, cytotoxicity was evaluated by a CFSE-based assay [79]. The percentage specific lysis was calculated as follows: (number of remaining $CFSE^+$ cells after incubation of target cells alone – number of remaining $CFSE^+$ cells after incubation with effector cells / number of $CFSE^+$ cells after incubation of target cells alone) x100.

4.6 Acknowledgments

We are grateful to the staff of the following core facilities at the Institute for Research in Immunology and Cancer for their outstanding support: Animal facility, Bioinformatics, Flow cytometry, and Proteomics. We thank Dr. R. Glas for kindly providing us with WT EL4, C4.4-25^{-} and E50.16^{+} cell lines. This work was supported by funds from the Canadian Cancer Society and the Terry Fox Foundation through the National Cancer Institute of Canada. MHF and EC are supported by training grants from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research, respectively. CP and PT hold Canada Research Chairs in Immunobiology, and Proteomics and Bioanalytical Spectrometry, respectively. The authors have no conflicting financial interests.

4.7 References

1. Rammensee, H.G., Falk, K., and Rotzschke, O., *Peptides naturally presented by MHC class I molecules.* Annu Rev Immunol, 1993. **11**: p. 213-44.

2. Yewdell, J.W., Reits, E., and Neefjes, J., *Making sense of mass destruction: quantitating MHC class I antigen presentation.* Nat Rev Immunol, 2003. **3**(12): p. 952-61.

3. Rock, K.L., York, I.A., Saric, T., and Goldberg, A.L., *Protein degradation and the generation of MHC class I-presented peptides.* Adv Immunol, 2002. **80**: p. 1-70.

4. Heemels, M.T. and Ploegh, H., *Generation, translocation, and presentation of MHC class I-restricted peptides.* Annu Rev Biochem, 1995. **64**: p. 463-491.

5. Pamer, E. and Cresswell, P., *Mechanisms of MHC class I--restricted antigen processing.* Annu Rev Immunol, 1998. **16**: p. 323-58.

6. Hammer, G.E., Kanaseki, T., and Shastri, N., *The final touches make perfect the peptide-MHC class I repertoire.* Immunity., 2007/4. **26**(4): p. 397-406.

7. Zhang, Y. and Williams, D.B., *Assembly of MHC class I molecules within the endoplasmic reticulum.* Immunol Res, 2006. **35**(1-2): p. 151-62.

8. Wong, P. and Pamer, E.G., *CD8 T cell responses to infectious pathogens.* Annu Rev Immunol, 2003. **21**: p. 29-70.

9. Weinzierl, A.O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.G., and Stevanovic, S., *Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface.* Mol Cell Proteomics, 2007. **6**(1): p. 102-13.

10. Starr, T.K., Jameson, S.C., and Hogquist, K.A., *Positive and negative T cell selection.* Annu Rev Immunol, 2003. **21**: p. 139-176.

11. Huseby, E.S., White, J., Crawford, F., Vass, T., Becker, D., Pinilla, C., Marrack, P., and Kappler, J.W., *How the T cell repertoire becomes peptide and MHC specific.* Cell, 2005. **122**(2): p. 247-260.

12. Marrack, P. and Kappler, J., *Control of T cell viability.* Annu Rev Immunol, 2004. **22**: p. 765-787.

13. Anikeeva, N., Lebedeva, T., Clapp, A.R., Goldman, E.R., Dustin, M.L., Mattoussi, H., and Sykulev, Y., *Quantum dot/peptide-MHC biosensors reveal strong CD8-dependent cooperation between self and viral antigens that augment the T cell response.* Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16846-16851.

14. Dunn, G.P., Old, L.J., and Schreiber, R.D., *The immunobiology of cancer immunosurveillance and immunoediting.* Immunity, 2004. **21**(2): p. 137-148.

15. Slev, P.R., Nelson, A.C., and Potts, W.K., *Sensory neurons with MHC-like peptide binding properties: disease consequences.* Curr Opin Immunol, 2006. **18**(5): p. 608-616.

16. Perreault, C., Decary, F., Brochu, S., Gyger, M., Belanger, R., and Roy, D., *Minor histocompatibility antigens.* Blood, 1990. **76**(7): p. 1269-80.

17. Liblau, R.S., Wong, F.S., Mars, L.T., and Santamaria, P., *Autoreactive CD8 T cells in organ-specific autoimmunity: emerging targets for therapeutic intervention.* Immunity, 2002. **17**(7): p. 1-6.

18. Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., and Engelhard, V.H., *Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry.* Science, 1992. **255**(5049): p. 1261-3.

19. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., and Rammensee, H.G., *Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules.* Nature, 1991. **351**(6324): p. 290-6.

20. Huczko, E.L., Bodnar, W.M., Benjamin, D., Sakaguchi, K., Zhu, N.Z., Shabanowitz, J., Henderson, R.A., Appella, E., Hunt, D.F., and Engelhard, V.H., *Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling.* J Immunol, 1993. **151**(5): p. 2572-87.

21. McBride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A., Bergeron, J., and Perreault, C., *The model B6(dom1) minor histocompatibility antigen is encoded by a mouse homolog of the yeast STT3 gene.* Immunogenetics, 2002. **54**(8): p. 562-9.

22. Admon, A., Barnea, E., and Ziv, T., *Tumor antigens and proteomics from the point of view of the major histocompatibility complex peptides.* Mol Cell Proteomics, 2003. **2**(6): p. 388-98.

23. Gebreselassie, D., Spiegel, H., and Vukmanovic, S., *Sampling of major histocompatibility complex class I-associated peptidome suggests relatively looser*

*global association of HLA-B*5101 with peptides.* Hum Immunol, 2006. **67**(11): p. 894-906.

24. Barnea, E., Beer, I., Patoka, R., Ziv, T., Kessler, O., Tzehoval, E., Eisenbach, L., Zavazava, N., and Admon, A., *Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens.* Eur J Immunol, 2002. **32**(1): p. 213-222.

25. Hickman, H.D., Luis, A.D., Buchli, R., Few, S.R., Sathiamurthy, M., VanGundy, R.S., Giberson, C.F., and Hildebrand, W.H., *Toward a definition of self: proteomic evaluation of the class I peptide repertoire.* J Immunol, 2004. **172**(5): p. 2944-52.

26. Hickman, H.D., Luis, A.D., Bardet, W., Buchli, R., Battson, C.L., Shearer, M.H., Jackson, K.W., Kennedy, R.C., and Hildebrand, W.H., *Cutting edge: class I presentation of host peptides following HIV infection.* J Immunol, 2003. **171**(1): p. 22-6.

27. Lemmel, C., Weik, S., Eberle, U., Dengjel, J., Kratt, T., Becker, H.D., Rammensee, H.G., and Stevanovic, S., *Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labeling.* Nat Biotechnol., 2004. **22**(4): p. 450-454.

28. Milner, E., Barnea, E., Beer, I., and Admon, A., *The turnover kinetics of major histocompatibility complex peptides of human cancer cells.* Mol Cell Proteomics, 2006. **5**(2): p. 357-65.

29. Meiring, H.D., Soethout, E.C., Poelen, M.C., Mooibroek, D., Hoogerbrugge, R., Timmermans, H., Boog, C.J., Heck, A.J., de Jong, A.P., and van Els, C.A., *Stable isotope tagging of epitopes: a highly selective strategy for the identification of major histocompatibility complex class I-associated peptides induced upon viral infection.* Mol Cell Proteomics, 2006. **5**(5): p. 902-13.

30. Mann, M., *Functional and quantitative proteomics using SILAC.* Nat Rev Mol Cell Biol, 2006. **7**(12): p. 952-8.

31. Ong, S.E., Foster, L.J., and Mann, M., *Mass spectrometric-based approaches in quantitative proteomics.* Methods, 2003. **29**(2): p. 124-30.

32. Caron, E., Charbonneau, R., Huppe, G., Brochu, S., and Perreault, C., *The structure and location of SIMP/STT3B account for its prominent imprint on the MHC I immunopeptidome.* Int Immunol, 2005. **17**(12): p. 1583-96.

33. Yewdell, J.W. and Nicchitta, C.V., *The DRiP hypothesis decennial: support, controversy, refinement and extension.* Trends Immunol, 2006. **27**(8): p. 368-73.

34. Benoist, C., Germain, R.N., and Mathis, D., *A plaidoyer for 'systems immunology'.* Immunol Rev, 2006. **210**: p. 229-234.

35. Fortier, M.H., Bonneil, E., Goodley, P., and Thibault, P., *Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs.* Anal Chem, 2005. **77**(6): p. 1631-40.

36. Kearney, P. and Thibault, P., *Bioinformatics meets proteomics--bridging the gap between mass spectrometry data analysis and cell biology.* J Bioinform Comput Biol, 2003. **1**(1): p. 183-200.

37. Storkus, W.J., Zeh, H.J., 3rd, Salter, R.D., and Lotze, M.T., *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.

38. Perreault, C., Jutras, J., Roy, D.C., Filep, J.G., and Brochu, S., *Identification of an immunodominant mouse minor histocompatibility antigen (MiHA). T cell response to a single dominant MiHA causes graft-versus-host disease.* J Clin Invest, 1996. **98**(3): p. 622-628.

39. Purcell, A.W., *Isolation and characterization of naturally processed MHC-bound peptides from the surface of antigen-presenting cells.* Methods Mol Biol, 2004. **251**: p. 291-306.

40. Glas, R., Sturmhofel, K., Hammerling, G.J., Karre, K., and Ljunggren, H.G., *Restoration of a tumorigenic phenotype by beta 2-microglobulin transfection to EL-4 mutant cells.* J Exp Med, 1992. **175**(3): p. 843-6.

41. Williams, D.B., Barber, B.H., Flavell, R.A., and Allen, H., *Role of beta 2-microglobulin in the intracellular transport and surface expression of murine class I histocompatibility molecules.* J Immunol, 1989. **142**(8): p. 2796-806.

42. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S., *SYFPEITHI: database for MHC ligands and peptide motifs.* Immunogenetics, 1999. **50**(3-4): p. 213-219.

43. Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D.C., Sidney, J., Bui, H.H., Grey, H., and Sette, A., *A consensus epitope prediction approach identifies the breadth of murine T CD8+-cell responses to vaccinia virus* Nat Biotechnol., 2006. **24**(7): p. 817-819.

44. Peters, B., Bui, H.H., Frankild, S., Nielson, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Fleri, W., Wilson, S.S., Sidney, J., Lund, O., Buus, S., and Sette, A., *A community resource benchmarking predictions of peptide binding to MHC-I molecules.* PLoS Comput Biol, 2006. **2**(6): p. e65.

45. Peters, B. and Sette, A., *Integrating epitope data into the emerging web of biomedical knowledge resources.* Nat Rev Immunol, 2007. **7**(6): p. 485-90.

46. Aldrich, C.J., DeCloux, A., Woods, A.S., Cotter, R.J., Soloski, M.J., and Forman, J., *Identification of a Tap-dependent leader peptide recognized by alloreactive T cells specific for a class Ib antigen.* Cell, 1994. **79**(4): p. 649-58.

47. Rotzschke, O., Falk, K., Stevanovic, S., Grahovac, B., Soloski, M.J., Jung, G., and Rammensee, H.G., *Qa-2 molecules are peptide receptors of higher stringency than ordinary class I molecules.* Nature, 1993. **361**(6413): p. 642-4.

48. Qian, S.B., Princiotta, M.F., Bennink, J.R., and Yewdell, J.W., *Characterization of rapidly degraded polypeptides in mammalian cells reveals a novel layer of nascent protein quality control.* J Biol Chem, 2006. **281**(1): p. 392-400.

49. Qian, S.B., Reits, E., Neefjes, J., Deslich, J.M., Bennink, J.R., and Yewdell, J.W., *Tight linkage between translation and MHC class I peptide ligand generation implies specialized antigen processing for defective ribosomal products.* J Immunol, 2006. **177**(1): p. 227-33.

50. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R., and Hogenesch, J.B., *A gene atlas of the mouse and human protein-encoding transcriptomes.* Proc Natl Acad Sci U S A, 2004. **101**(16): p. 6062-7.

51. Zhang, W., Morris, Q.D., Chang, R., Shai, O., Bakowski, M.A., Mitsakakis, N., Mohammad, N., Robinson, M.D., Zirngibl, R., Somogyi, E., Laurin, N., Eftekharpour, E., Sat, E., Grigull, J., Pan, Q., Peng, W.T., Krogan, N., Greenblatt, J., Fehlings, M., van der Kooy, D., Aubin, J., Bruneau, B.G., Rossant, J., Blencowe, B.J., Frey, B.J., and Hughes, T.R., *The functional landscape of mouse gene expression.* J Biol, 2004. **3**(5): p. 21.

52. Virtaneva, K., Wright, F.A., Tanner, S.M., Yuan, B., Lemon, W.J., Caligiuri, M.A., Bloomfield, C.D., de La Chapelle, A., and Krahe, R., *Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics.* Proc Natl Acad Sci U S A, 2001. **98**(3): p. 1124-9.

53. Dorn, T., Kuhn, U., Bungartz, G., Stiller, S., Bauer, M., Ellwart, J., Peters, T., Scharffetter-Kochanek, K., Semmrich, M., Laschinger, M., Holzmann, B., Klinkert, W.E., Straten, P.T., Kollgaard, T., Sixt, M., and Brakebusch, C., *RhoH is important for positive thymocyte selection and T-cell receptor signaling.* Blood, 2007. **109**(6): p. 2346-55.

54. Plotkin, J., Prockop, S.E., Lepique, A., and Petrie, H.T., *Critical role for CXCR4 signaling in progenitor localization and T cell differentiation in the postnatal thymus.* J Immunol, 2003. **171**(9): p. 4521-7.

55. Malissen, M., Gillet, A., Ardouin, L., Bouvier, G., Trucy, J., Ferrier, P., Vivier, E., and Malissen, B., *Altered T cell development in mice with a targeted mutation of the CD3-epsilon gene.* Embo J, 1995. **14**(19): p. 4641-53.

56. Richter, J.D. and Sonenberg, N., *Regulation of cap-dependent translation by eIF4E inhibitory proteins.* Nature, 2005. **433**(7025): p. 477-480.

57. Nakayama, K.I. and Nakayama, K., *Ubiquitin ligases: cell-cycle control and cancer.* Nat Rev Cancer, 2006. **6**(5): p. 369-81.

58. Siegel, J.N., Turner, C.A., Klinman, D.M., Wilkinson, M., Steinberg, A.D., MacLeod, C.L., Paul, W.E., Davis, M.M., and Cohen, D.I., *Sequence analysis and expression of an X-linked, lymphocyte-regulated gene family (XLR).* J Exp Med, 1987. **166**(6): p. 1702-15.

59. Hennessy, B.T., Smith, D.L., Ram, P.T., Lu, Y., and Mills, G.B., *Exploiting the PI3K/AKT pathway for cancer drug discovery.* Nat Rev Drug Discov, 2005. **4**(12): p. 988-1004.

60. Mamane, Y., Petroulakis, E., LeBacquer, O., and Sonenberg, N., *mTOR, translation initiation and cancer.* Oncogene, 2006/10/16. **25**(48): p. 6416-6422.

61. Singh-Jasuja, H., Emmerich, N.P., and Rammensee, H.G., *The Tubingen approach: identification, selection, and validation of tumor-associated HLA peptides for cancer therapy.* Cancer Immunol Immunother., 2004. **53**(3): p. 187-195.

62. Purcell, A.W., McCluskey, J., and Rossjohn, J., *More than one reason to rethink the use of peptides in vaccine design.* Nat Rev Drug Discov, 2007. **6**(5): p. 404-14.

63. Suri, A., Walters, J.J., Levisetti, M.G., Gross, M.L., and Unanue, E.R., *Identification of naturally processed peptides bound to the class I MHC molecule H-2Kd of normal and TAP-deficient cells.* Eur J Immunol, 2006. **36**(3): p. 544-57.

64. Gleimer, M. and Parham, P., *Stress management: MHC class I and class I-like molecules as reporters of cellular stress.* Immunity, 2003. **19**(4): p. 469-77.

65. Glickman, M.H. and Ciechanover, A., *The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction.* Physiol Rev., 2002. **82**(2): p. 373-428.

66. Matos, M., Park, R., Mathis, D., and Benoist, C., *Progression to islet destruction in a cyclophosphamide-induced transgenic model: a microarray overview.* Diabetes, 2004. **53**(9): p. 2310-21.

67. Ruggero, D. and Pandolfi, P.P., *Does the ribosome translate cancer?* Nat Rev Cancer, 2003. **3**(3): p. 179-92.

68. Nalepa, G., Rolfe, M., and Harper, J.W., *Drug discovery in the ubiquitin-proteasome system.* Nat Rev Drug Discov, 2006. **5**(7): p. 596-613.

69. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., and Golub, T.R., *MicroRNA expression profiles classify human cancers.* Nature, 2005. **435**(7043): p. 834-8.

70. Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R., *Chromosomal instability and tumors promoted by DNA hypomethylation.* Science, 2003. **300**(5618): p. 455.

71. Yoshimura, Y., Yadav, R., Christianson, G.J., Ajayi, W.U., Roopenian, D.C., and Joyce, S., *Duration of alloantigen presentation and avidity of T cell antigen recognition correlate with immunodominance of CTL response to minor histocompatibility antigens.* J Immunol, 2004. **172**(11): p. 6666-74.

72. Louis, I., Dulude, G., Corneau, S., Brochu, S., Boileau, C., Meunier, C., Côté, C., Labrecque, N., and Perreault, C., *Changes in the lymph node microenvironment induced by Oncostatin M.* Blood, 2003. **102**(4): p. 1397-1404.

73. Meunier, M.C., Roy-Proulx, G., Labrecque, N., and Perreault, C., *Tissue distribution of target antigen has a decisive influence on the outcome of adoptive cancer immunotherapy.* Blood, 2003. **101**(2): p. 766-770.

74. Pion, S., Fontaine, P., Baron, C., Gyger, M., and Perreault, C., *Immunodominant minor histocompatibility antigens expressed by mouse leukemic cells can serve as effective targets for T cell immunotherapy.* J Clin Invest, 1995. **95**(4): p. 1561-8.

75. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C., and Lempicki, R.A., *Identifying biological themes within lists of genes with EASE.* Genome Biol, 2003. **4**(10): p. R70.

76. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

77. Pion, S., Fontaine, P., Desaulniers, M., Jutras, J., Filep, J.G., and Perreault, C., *On the mechanisms of immunodominance in cytotoxic T lymphocyte responses to minor histocompatibility antigens.* Eur J Immunol, 1997. **27**(2): p. 421-30.

78. Galea-Lauri, J., Wells, J.W., Darling, D., Harrison, P., and Farzaneh, F., *Strategies for antigen choice and priming of dendritic cells influence the polarization and*

efficacy of antitumor T-cell responses in dendritic cell-based cancer vaccination. Cancer Immunol Immunother, 2004. **53**(11): p. 963-77.

79. Jedema, I., van der Werff, N.M., Barge, R.M., Willemze, R., and Falkenburg, J.H., *New CFSE-based assay to determine susceptibility to lysis by cytotoxic T cells of leukemic precursor cells within a heterogeneous target cell population.* Blood, 2004. **103**(7): p. 2677-82.

5. The mTOR signaling pathway and its influence on the MHC class I peptide repertoire

Marie-Hélène Fortier, Etienne Caron, Claude Perreault, and Pierre Thibault

PNAS, 2008. (en voie d'être soumis)

5.1 Abstract

Proteins associated with the mammalian target of rapamycin (mTOR) signaling pathway are often overexpressed or mutated in a large number of cancers. This pathway plays an essential role in the regulation of processes such as ribosome biogenesis and protein translation, which are critical for cell growth, proliferation and differentiation. In this work, we studied the influence of mTOR on the MHC class I peptide (MIP) repertoire for two major reasons: i) the tremendous importance of this pathway in oncogenic processes ii) its role in the control of protein synthesis, which is at the origin of the generation of the MIP repertoire. To achieve this goal, we developed a high-throughput MS-based quantitative approach to monitor the interplay between the MIP repertoire and the mTOR signaling pathway. The kinetic profiles of MHC I peptides from control and rapamycin-treated EL4 cells indicated a progressive increase in abundance of 70% of the MIP repertoire, which correlates with an increase level of cell surface expression of MHC I molecules. Out of this number, 17% of peptides were differentially expressed following prolonged mTOR inhibition, and encoded a large proportion of source proteins associated to cell cycle progression and cellular proliferation. Comparison of the abundance profiles of MHC I peptides and the steady-state level of their corresponding source proteins led to the conclusion that the MIP repertoire does not reflect the proteome of the cell. Rather, these results suggest that defective *de novo* protein synthesis and/or protein degradation rates could influence more significantly the MIP repertoire following mTOR inhibition. We conclude that targeting mTOR by rapamycin drastically influence the composition of the MIP repertoire by enhancing antigenic presentation. This led us to propose that modulation of the mTOR signaling pathway might affect immunosurveillance mechanisms.

5.2 Introduction

MHC class I molecules present foreign and self-peptides (MIP repertoire) at the cell surface that can be recognized by the immune system [1]. Most of these peptides are generated from source proteins that undergo proteasome degradation [2]. The proteasome products are translocated in the endoplasmic reticulum (ER) via the transporter associated with antigenic processing (TAP) complex, where they are trimmed by ER aminopeptidase and incorporated in MHC I proteins by the peptide-loading complex (PLC) before presentation to the cell surface [3-6]. Important intracellular changes taking place in infected or transformed cells can then be reported to the immune system via the MIP repertoire to induce a CD8 T cell response [7, 8]. For example, stressful condition such as γ-irradiation can change the composition of the MIP repertoire by inducing the presentation of radiation-specific peptides [9]. Moreover, comparison of the MIP repertoire of normal and neoplastic thymocytes revealed that half of MHC I-associated peptides differentially expressed derived from proteins involved in neoplastic transformation [10].

The genesis of the MIP repertoire is mainly controlled by cellular processes such as protein translation and degradation [11, 12]. Therefore, the MIP repertoire might give unique insights into the mechanisms of cellular stress response and tumorigenesis. Of particular interest is the signaling pathways activated by the mammalian target of rapamycin (mTOR), known to be altered in different types of human cancers [13, 14]. mTOR plays an important role in integrating extracellular signals with intracellular amino acid pool and energy metabolism to affect mRNA translation, ribosomal biogenesis, cell cycle progression and autophagy, which are critical for cell growth, proliferation and differentiation [14-16]. In addition, hyperactivation of the mTOR signaling pathway is a key feature in many type of cancer because it confers a significant advantage for tumor development and progression [13]. Previous investigations on the presentation of MHC class I peptides from neoplastic thymocytes led to the suggestion that the mTOR pathway could influence the MIP repertoire since a significant proportion of the corresponding source proteins were derived from molecules implicated in the PI3K-AKT-mTOR pathway [10]. In the present study, we evaluated the extent to which the mTOR signaling pathway can affect the MIP repertoire by measuring the dynamic changes in the composition of the

MIP repertoire following inhibition of mTOR with rapamycin. The availability of efficient quantitative proteomics tools to comprehensively mine the MIP repertoire lead to a high definition of molecular pathways providing meaningful insights into the genesis and the composition of MHC class I peptides in response to cellular stress.

5.3 Results

5.3.1 mTOR inhibition affects cell size, cell growth and apoptosis in EL4 cells

Hyperactivation of mTOR has been reported in a large number of cancer cells [13], and specific inhibition of its activity by rapamycin is currently investigated in numerous clinical trials [17]. However, the effects of rapamycin on protein degradation and MHC I peptide presentation remains largely unknown. Accordingly, we investigated the effect of mTOR inhibition on the MIP repertoire of EL4 neoplastic thymocytes derived from a C57BL/6 mouse.

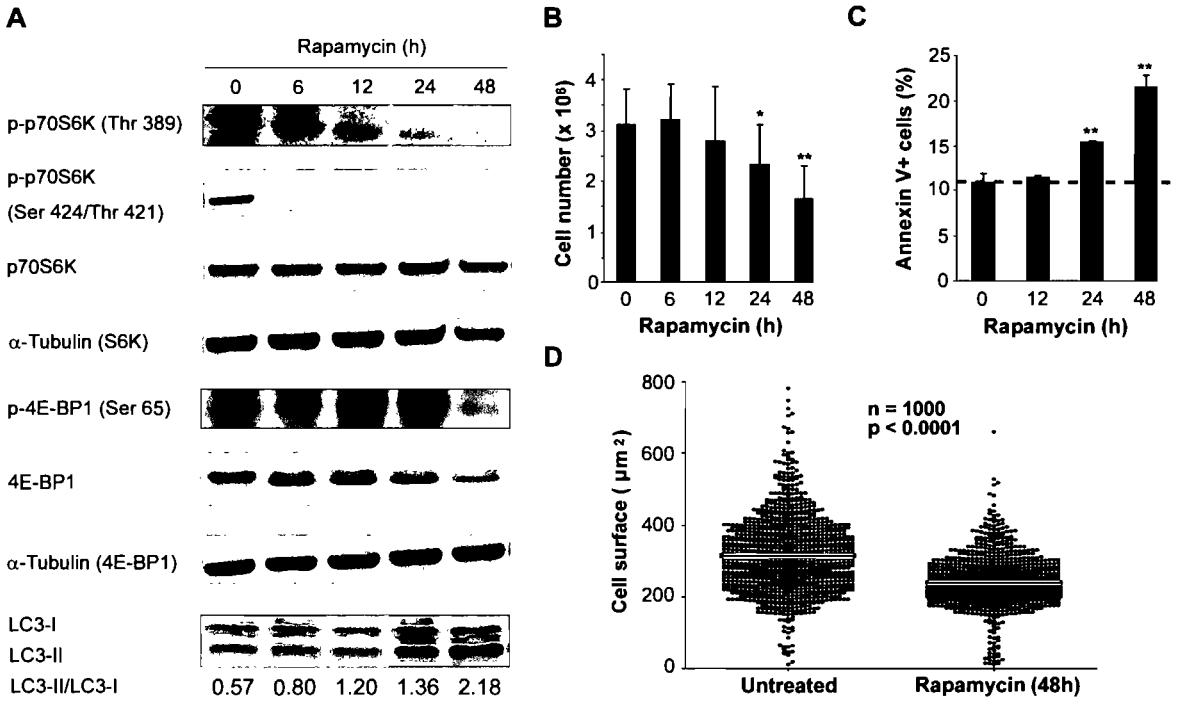

Figure 5.1: mTOR inhibition in EL4 cells. (A) Western blot kinetic showing dephosphorylation of p70S6K (Thr^{389}, Thr^{421}/Ser^{424}) and 4E-BP1 (Ser^{65}) following rapamycin exposure. Western blot kinetic on autophagy marker LC3. Effect of mTOR inhibition on (B) cell growth, (C) apoptosis and (D) cell size (n=4; mean ± SD). Estimation of apoptotic cells were based on staining with PE-conjugated annexin V antibody. Cell number and cell size were measured by light microscopy. *P<0.05; **P<0.005.

We first confirmed mTOR inhibition by monitoring the dephosphorylation of mTOR downstream effectors p70S6K and 4E-BP1 following rapamycin treatment [18].

Dephosphorylation of p70S6K (Thr^{389}, Thr^{421}/Ser^{424}) and 4E-BP1 (Ser^{65}) were observed in rapamycin-treated EL4 cells up to 48 h (Figure 5.1A). We also observed a dephosphorylation of Ser^{65} from 4E-BP1 while the non phosphorylated 4E-BP1 remained stably expressed until 24 h following rapamycin treatment. The decrease in the steady-state level of 4E-BP1 protein observed at 48 h could be possibly explained as a consequence of the p70S6K dependent feedback mechanism becoming more important following prolonged rapamycin treatment.

Treatment with rapamycin can lead to noticeable phenotypical changes including inhibition of cell proliferation, cell size reduction, autophagy and apoptosis [15, 19]. We monitored induction of autophagy in EL4 cells using the LC3 marker by following conversion of LC3-I into LC3-II [18]. An increase in LC3-II:LC3-I ratio was observed consistently over the time period examined (Figure 5.1A). We also noticed a two-fold reduction in cellular proliferation (Figure 5.1B) with a two-fold increase in the number of annexin V positive-apoptotic cells (Figure 5.1C). This was accompanied by a gradual decrease in cell volume and size distribution over the 48 h period following rapamycin treatment (Figure 5.1D).

5.3.2 Identification and dynamic profiling of the MIP repertoire from EL4 cells

To study the influence of the mTOR pathway on the MIP repertoire, we treated 500 millions EL4 cells for each rapamycin treatment and periodically extracted MHC I-associated peptides from the cell surface using mild acid elution (MAE) [10]. Samples obtained from MAE were then fractioned by strong cation exchange chromatography (SCX) and analyzed using nanoLC-MS on an Orbitrap mass spectrometer (Figure 5.2). To monitor dynamic changes in the composition of the MIP repertoire following mTOR inhibition, we used a label-free quantitative approach as described previously [10] (Figure 5.2, see materials and methods for more details). In total, we identified 593 unique MHC I peptide structures (Table AIII.I). The uniqueness of the MHC I peptides was also confirmed using a β_2m-deficient control sample [10]. Classification of these peptides according to their respective allele showed the following MHC I motifs: 301 $H2K^b$-, 253 $H2D^b$-, 38 Qa2-, and 1 Qa1-associated peptides (Table AIII.I). GO terms enrichment analysis was performed on the group of genes encoding EL4-MHC I peptides [20],

considering only peptide sequences assigned to unique genes (521). We found that the MIP repertoire of EL4 cells contains a 2-to-4-fold enrichment in proteins related to regulation of cell cycle, ribosome biogenesis and assembly, translation and response to DNA damage stimulus compared to the overall mouse genome (Figure 5.3). Interestingly, GO terms enriched in neoplastic thymocytes are different from those that were significantly enriched in primary normal thymocytes [10]. These results support the concept that metabolic changes that take place within the cell are reflected in the composition of the MIP repertoire [21].

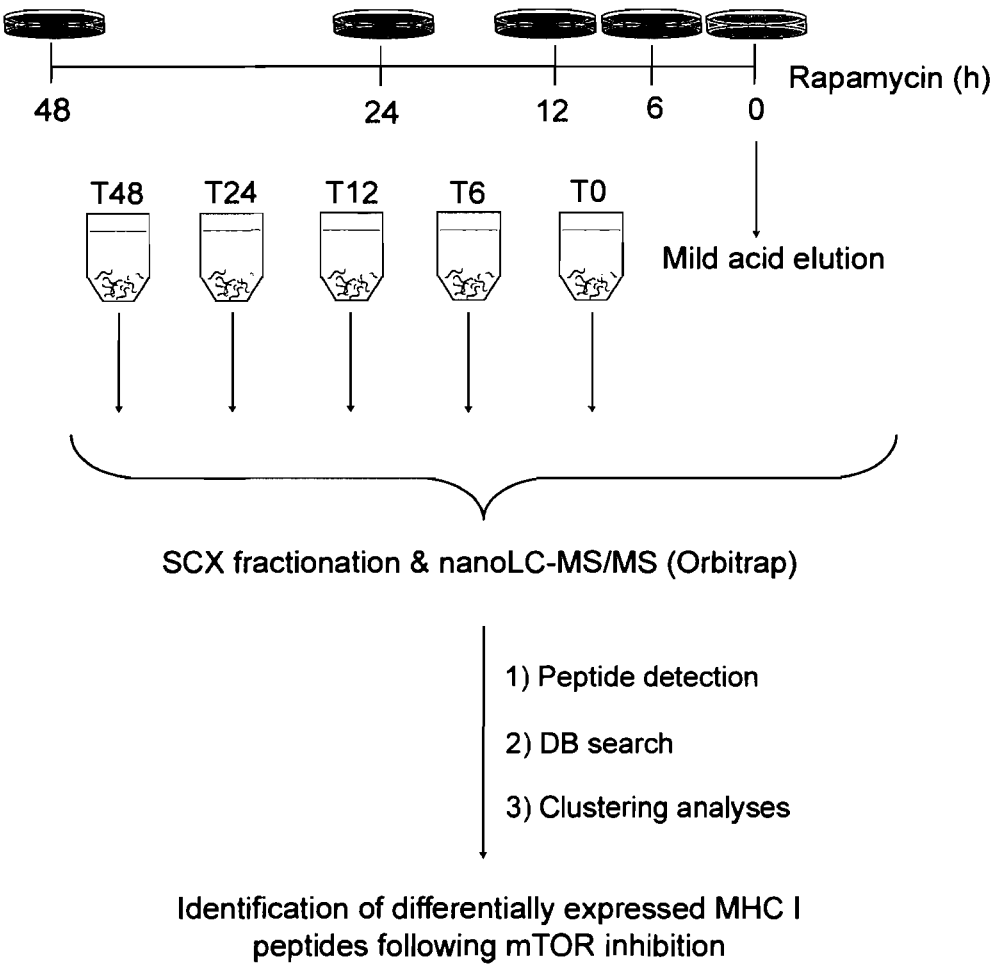

Figure 5.2: Experimental approach for the dynamic quantification of the MIP repertoire of EL4 cells. MHC I peptides from control and rapamycin-treated EL4 cells (20 ng/mL, T = 6, 12, 24, 48 hrs) were eluted from cell surface using MAE. Peptides were fractioned by SCX chromatography and analyzed using nanoLC-MS (three biological replicates for each time points). To identify differentially expressed MHC I peptides candidates following mTOR inhibition, peptide detection algorithms and segmentation analyses were used to compare unlabeled peptide populations.

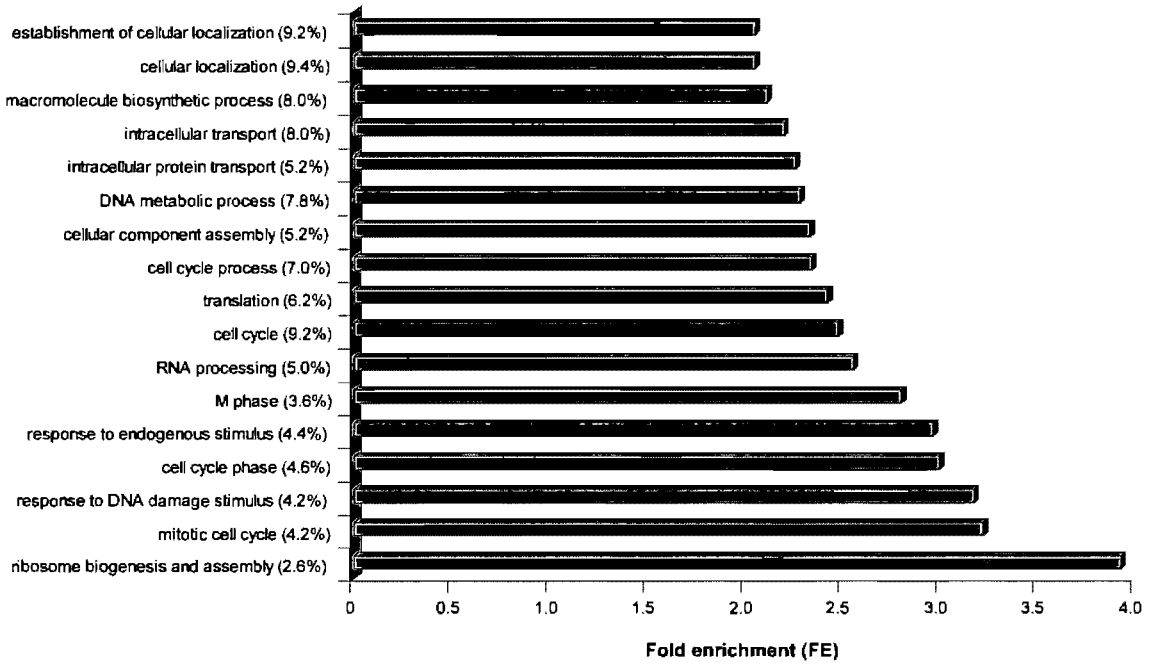

Figure 5.3: GO terms enrichment analysis of genes coding EL4-MHC I peptides. Biological processes in which EL4-MHC I peptide source proteins take part are listed according to their fold enrichment relative to the mouse genome. Values in parentheses represent percentages of peptide source proteins falling in each GO category. Exact P values and Benjamini values were < 0.05 for each listed GO term.

5.3.3 Global overexpression of the MIP repertoire of neoplastic EL4 thymocytes following prolonged mTOR inhibition

To profile the global changes in the abundance and identify groups of MHC I peptides showing specific trends in response to rapamycin treatment, we first clustered our data according to their source alleles. From a total of 593 MHC I peptides, we identified 425 peptides representing four predominant alleles that were reproducibly detected across all 5 time points (Table AIII.II). The heat map representation of the temporal distribution of these peptides identified two major groups: (i) MHC I-associated peptides showing a progressive upregulation over the rapamycin incubation period examined, and (ii) MHC I-associated peptides with unchanged expression during rapamycin treatment (Figure 5.4A).

Remarkably, these results show that ~ 70% of the MIP repertoire of EL4 cells is affected by mTOR inhibition (Figure 5.4B). The group of overexpressed peptides was predominately represented by $H2K^b$ (168) and $H2D^b$ molecules (119) compare to non-classical Qa1 and Qa2 molecules (6). To determine whether this overexpression was significant, we compared the extent of changes from early (T6-to-T0 ratios) to long-term exposure to rapamycin (T48-to-T0 ratios) for all 425 MHC I peptide clusters according to their source alleles (Figure 5.4C). We found a significant overexpression of $H2K^b$- and $H2D^b$-associated peptides ($P < 0.001$) but not for Qa2 peptides ($P = 0.06$) following prolonged mTOR inhibition using a Wilcoxon rank sum test with continuity correction. These results suggest that changes in the MHC I presentation pathway in response to rapamycin might have a more significant influence on peptides from specific alleles such as that of $H2K^b$ in the present situation. Interestingly, it was reported previously that $H2K^b$ molecules are exported more rapidly than $H2D^b$ molecules [22].

We next sought to determine whether the increase proportion of MHC I peptides following prolonged rapamycin treatment correlates with MHC I expression level of their respective source alleles. A 30% and 60% increase in abundance were observed for $H2K^b$ and $H2D^b$ molecules 48 h following rapamycin, respectively (Figure 5.5A). Moreover, we also found an increase of 30% in β2-microglobulin (β_2m) expression, a molecule essential for the formation of stable peptide-MHC I complexes. Flow cytometry analysis revealed that expression of Qa1 and Qa2 were decreased by 20% following prolonged rapamycin treatment (Figure 5.5A). These results may explain why variation in expression was not observed for most peptides derived from MHC Ib molecules (Figure 5.4A, B and C). Lymphocytes express two forms of Qa-2 molecules: Glycosylphosphatidylinositol (GPI)-modified, membrane-attached antigens and soluble molecules [23]. We hypothesized that rapamycin may diminish the activity of molecules implicated in the GPI-anchoring and/or modify the splicing of Qa-2. In our case, because most of the peptides overexpressed following mTOR inhibition derived from classical MHC Ia alleles, we decided to focus on peptide pools presented by $H2D^b$ and $H2K^b$ molecules. Together, these results suggest that rapamycin increases abundance of MHC Ia allelic products in EL4 cells, which correlates with the global overexpression observed for $H2D^b$- and $H2K^b$-associated peptides.

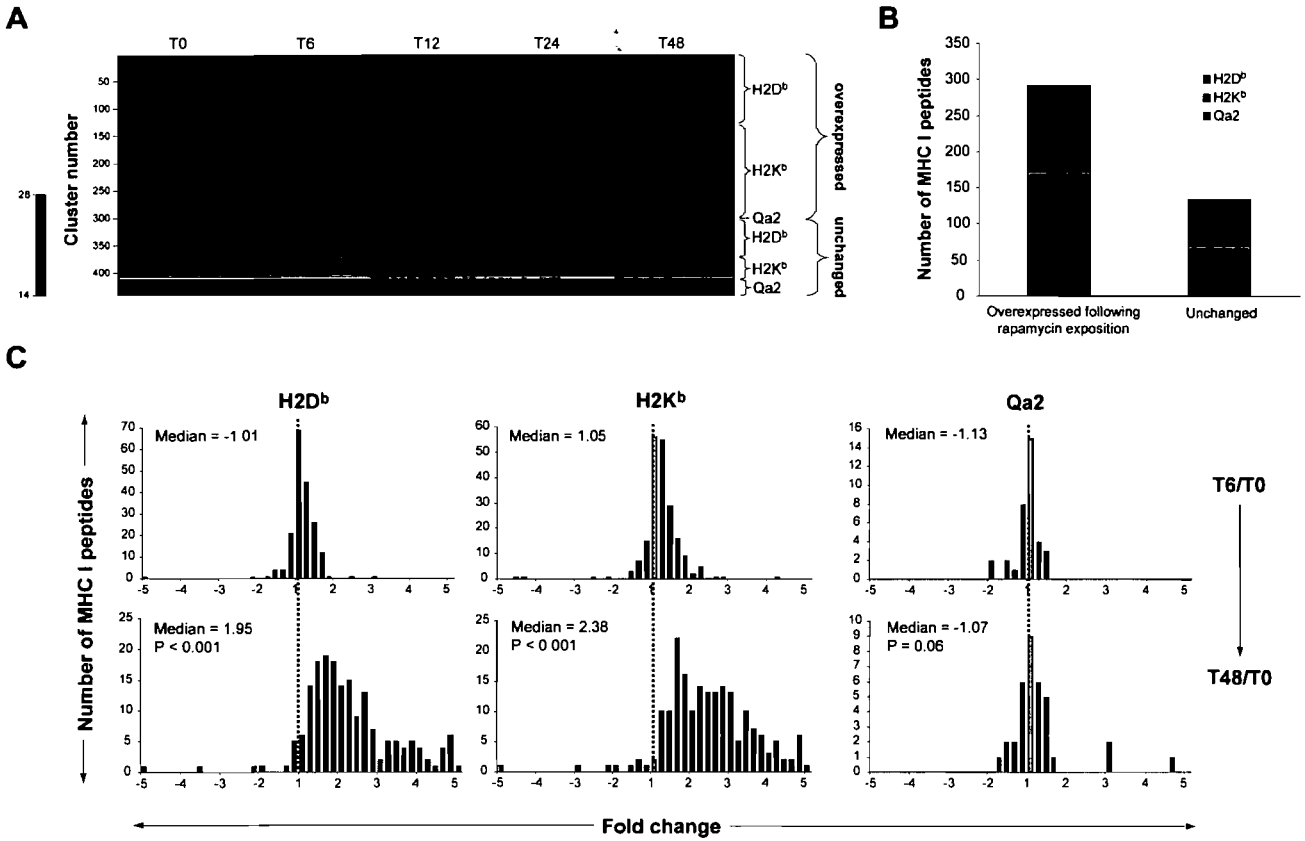

Figure 5.4: Kinetic profiles of the MIP repertoire of EL4 cells following mTOR inhibition. (A) Heat map representation shows differential expression between the different rapamycin incubation period (T = 0, 6, 12, 24, 48 h) where each horizontal line corresponds to a unique MHC I peptide clusters (n=425). A logarithmic scale distinguishes between peptides that are expressed at high (red) or low (green) level. (B) Distributions for the two clustering groups (defined in A) according to their peptide source alleles. (C) Fold change distributions comparison between T6/T0 and T48/T0 ratios for the different peptide source alleles (n_{H2Db} = 186, n_{H2Kb} = 205, n_{Qa2} = 34). Frequency distributions were plotted using a bin increment of 0.2. P values were calculated using the Wilcoxon rank sum test with continuity correction.

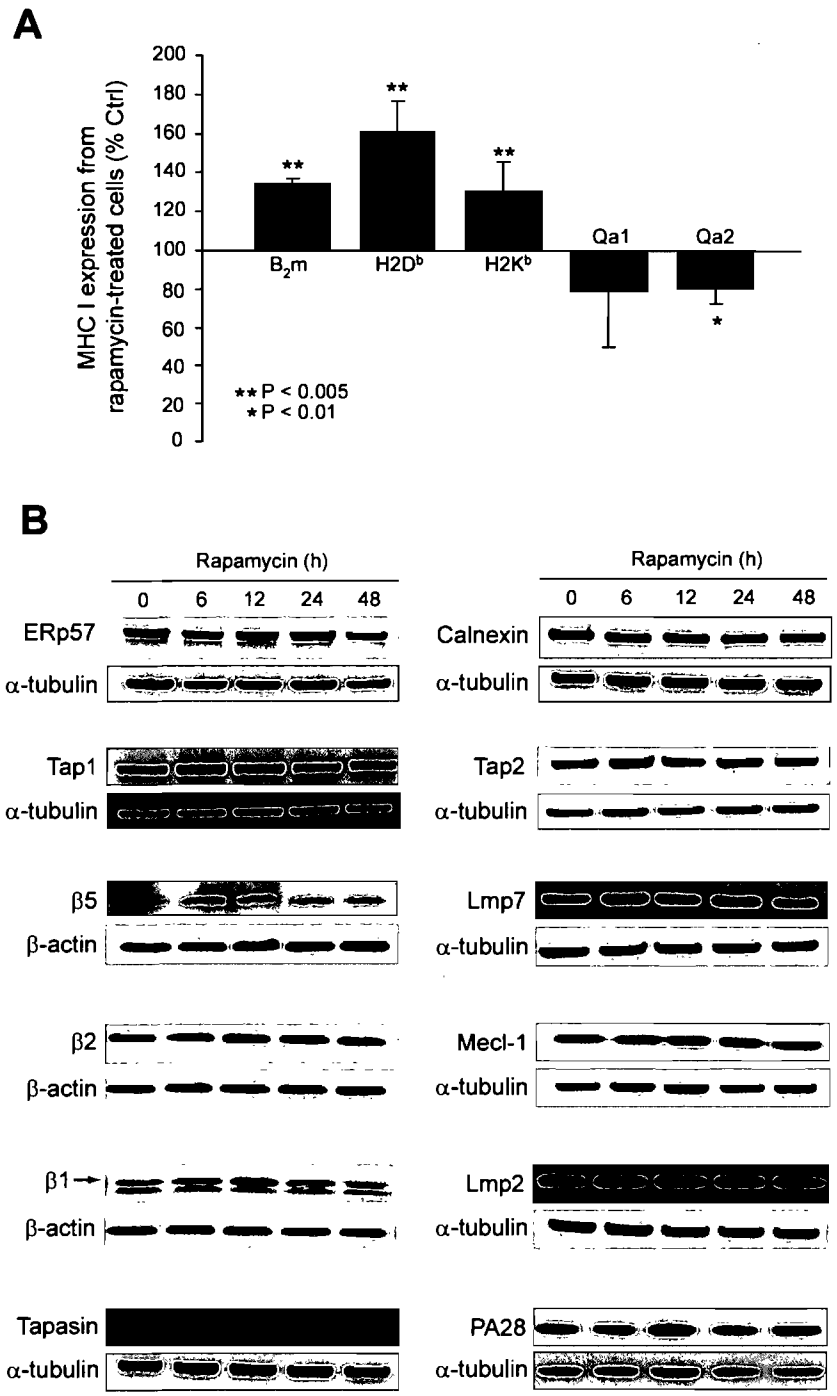

Figure 5.5: Effect of mTOR inhibition on MHC I presentation and expression levels of proteins involved in the MHC I processing machinery. (A) Comparison of cell surface MHC I expression between rapamycin-treated (T48) and control (T0) EL4 cells (n=4; mean ± SD). Cells were stained with antibodies against β_2m, H2D^b, H2K^b, Qa1 and Qa2. Expression levels are expressed in percentage compare to control (% Ctrl). *P<0.01; **P<0.005. (B) Kinetic profiles following mTOR inhibition on selected proteins involved in the MHC I processing machinery. Western blots were probed against the indicated proteins and loading control (α-tubulin or β-actin).

5.3.4 Influence of mTOR inhibition on the MHC I processing pathway of neoplastic thymocytes

MHC I peptide presentation is mainly controlled by the processing of source proteins along the antigen presentation pathway, where the peptide pool is the limiting factor in MHC I assembly [2]. Because ~ 70% of MHC I peptides were overexpressed following prolonged mTOR inhibition, we hypothesized that rapamycin might influence expression levels of key players involved in the MHC I processing pathway. The generation of peptides by the proteasome is one of the crucial steps required for antigen presentation. Cells have the ability to present two types of multisubunit proteases, which are defined as the standard proteasome (β1, β2, β5 subunits) and the immunoproteasome (β1i, β2i, β5i subunits) [24]. To monitor the effect of mTOR inhibition on the proteasome and immunoproteasome subunits, we compared kinetic profiles of targeted proteins using Western blot analysis (Figure 5.5B). These results indicate that both types of proteasomes are expressed in EL4 cells. We observed no significant changes for the proteasome subunits β2 and β5, the immunoproteasome subunits LMP 2 (β1i), Mecl-1 (β2i) and LMP 7 (β5i), and the proteasome activator PA28. In contrast, we observed a two-fold decrease of the subunit β1 compare to control EL4 cells following rapamycin treatment (48 h). Interestingly, decrease of the β1 subunit correlates with previously reported observations that mTOR inhibition decrease the proteasomal activity [25, 26].

Because efficient peptide-loading on MHC I molecules is a crucial step required for antigen presentation, we also monitored the influence of rapamycin on components of the peptide-loading complex (PLC). The PLC comprises many components including the chaperone calreticulin, TAP, ERp57 and tapasin [27]. Unchanged kinetic profiles were obtained for TAP1, TAP2 and ERp57 proteins (Figure 5.5B). The chaperone protein calnexin, which plays important roles in folding and oxidation of MHC I heavy chain and assembly with β_2m, was also unaffected by rapamycin. In contrast, a progressive overexpression of tapasin up to 5-fold at 24 and 48 h was observed following rapamycin treatment. Tapasin is an important component of the PLC as it assists the loading of antigenic peptides onto MHC I molecules and could play a significant role in peptide selection [28]. Moreover, tapasin is known to be expressed at low levels in a variety of

cancers, and its down-regulation leads to a reduction in the expression of MHC I molecules at the cell surface [29]. Our result suggests that overexpression of tapasin following rapamycin exposure may influence peptide composition at the cell surface.

5.3.5 The MIP repertoire of neoplastic thymocytes following mTOR inhibition

Our main objective was to evaluate the impact of the mTOR oncogenic pathway on the genesis and the composition of the MIP repertoire of neoplastic thymocytes. To this end, we compared the MIP repertoire of control and EL4 cells exposed to rapamycin for 48 h. First, we evaluated the reproducibility of the data obtained with our methodology by comparing MHC I peptides intensities across biological sets. We found that 95% of peptide ions showed variation of less than ± 2.3 fold in abundance across biological replicates (Figure 5.6). Peptides were considered differentially expressed when they showed a fold difference in abundance ≥ 2.5 with p-values < 0.05.

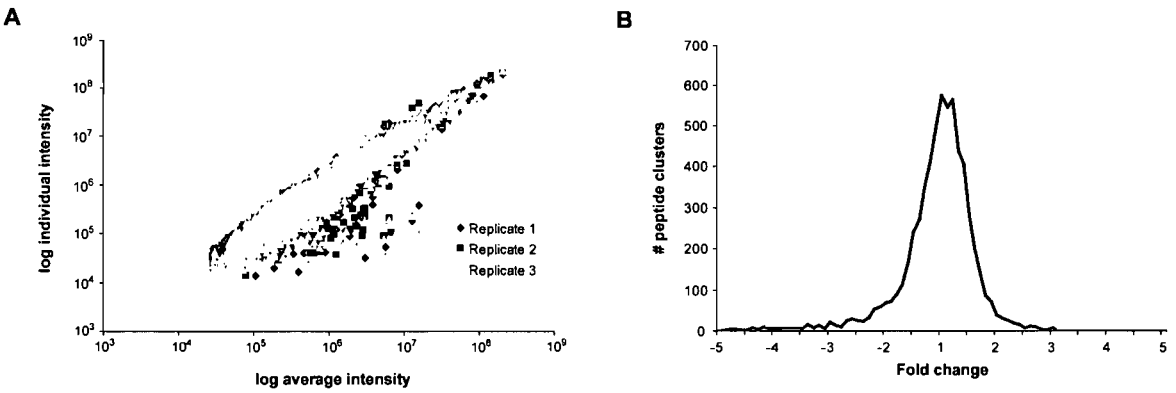

Figure 5.6: Reproducibility of intensity measurements from different biological cultures of EL4 cells. (A) Scatter plot of intensities for biological replicates at five time points (n=3). (B) Expression plot from different biological cultures (N= 6171 peptide clusters). Peptide ions (95%) show variation in intensity within ± 2.3 fold change.

Because we found a global overexpression of MHC I molecules following mTOR inhibition, fold change thresholds for $H2K^b$, $H2D^b$ and Qa2 peptides were normalized according to MHC I expression obtained by flow cytometry (Figure 5.5A). $H2K^b$, $H2D^b$ and Qa2 peptides were then considered to be differentially expressed when variation in abundance corresponds to fold change of ≤ -1.9 and ≥ 3.1 for $H2D^b$, ≤ -2.2 and ≥ 2.8 for

$H2K^b$, ≤ -2.7 and ≥ 2.3 for Qa2 (Figure 5.7A, blue dots). Accordingly, we found that 17% of MHC I peptides were differentially expressed following mTOR inhibition (Figure 5.7A, blue dots, Table AIII.II). We identified only one MHC I peptide that was significantly underexpressed following rapamycin treatment (Tmod1). We also identified 6 peptides only observed at the cell surface of EL4 cells following prolonged rapamycin treatment. Source genes encoding differentially expressed peptides are implicated in several biological processes such as cell cycle, chromatin related processes, regulation of transcription and translation, cell differentiation, signal transduction, metabolism, apoptosis, as well as implications in immune response, transport and cytoskeleton assembly (Table 5.1, Table AIII.III). From the 75 peptides differentially expressed, ~ 55% are derived from genes involved in cell cycle progression and/or cell proliferation and development (Table 5.1, Table AIII.III), which are important cellular processes controlled by the mTOR pathway [14, 16].

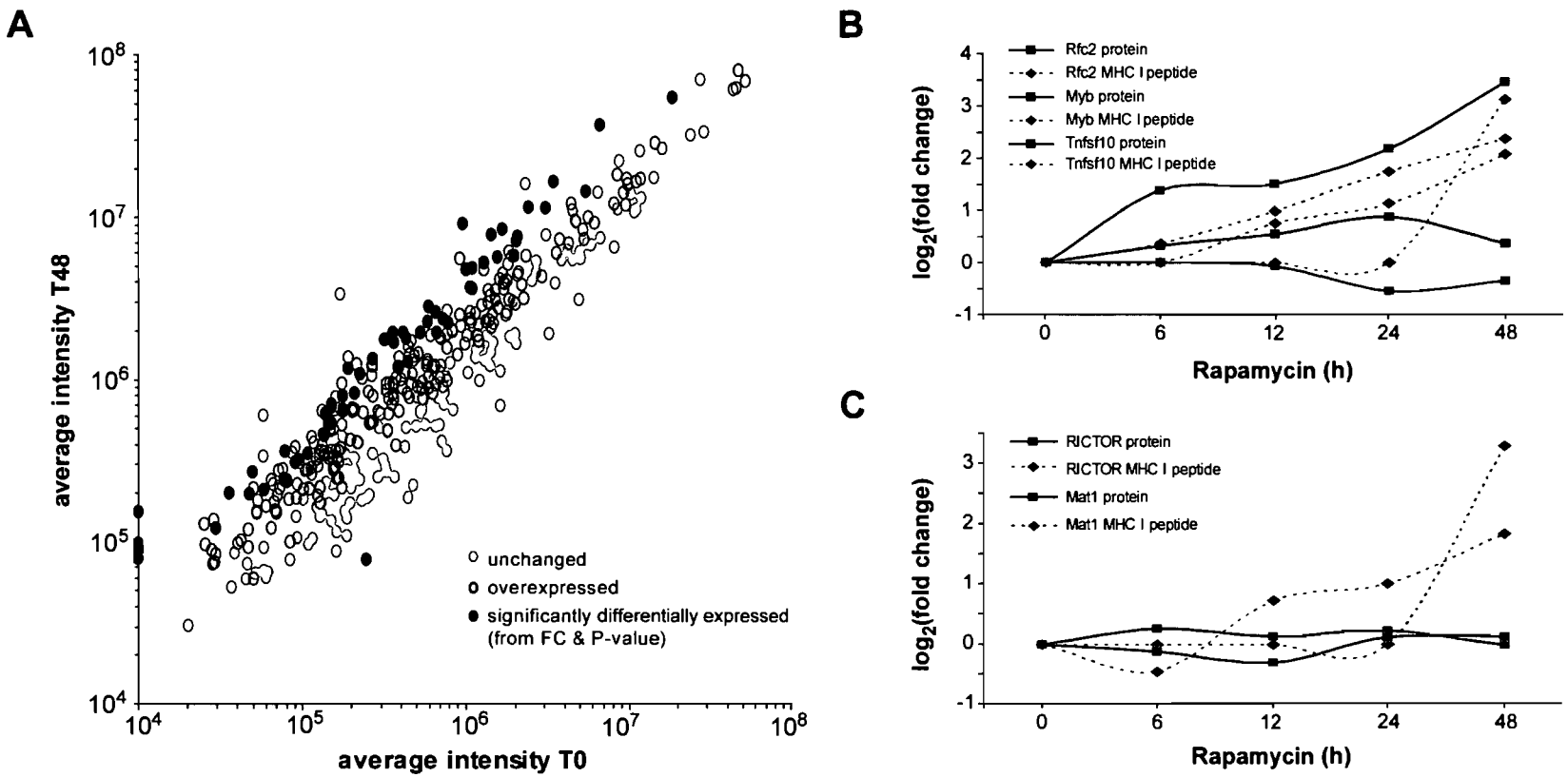

Figure 5.7: Relative quantification of differentially expressed MHC I peptides and expression levels of peptide-source proteins following mTOR inhibition in EL4 cells. (A) Scatter plot of intensities between Ctrl (T0) and prolonged rapamycin treatment (T48) for MHC I peptides reproducibly detected across biological replicates (n=3) and time courses (n=5). Peptides classified as overexpressed and significantly differentially expressed following prolonged mTOR inhibition (fold $change_{H2Db} \leq -1.9$ and ≥ 3.1, fold $change_{H2Kb} \leq -2.2$ and ≥ 2.8, fold $change_{Qa2} \leq -2.7$ and ≥ 2.3, $P \leq 0.05$) were highlighted in red and blue, respectively. (B and C) Kinetic profiles of peptide-source proteins and their corresponding overexpressed MHC I peptides. Band intensities from Western blot were measured and normalized against α-tubulin. Cell surface Tnfsf10 protein expression was evaluated by flow cytometry. Log base 2 plots were used to facilitate visualization of graphs.

Table 5.1: Examples of MHC I peptides differentially expressed on EL4 cells following prolonged mTOR inhibition.

Functional Classification	Gene Symbol	Gene ID	Function	Sequence	Rapa/ CTRL	P-value
Cell cycle	*Ccnf*	12449	Cyclin F, G2/M transition	SQAVNKQQI	6.0	0.02
	Lin37	75660	Member of the DREAM complex	STLIYRNM	3.4	0.05
	Mnat1	17420	Component of CAK, G1/S transition	YQKENKDVI	3.6	0.01
	Mphosph1	240641	M phase phosphoprotein 1	QSVAFTKL	4.1	0.003
	Tfdp1/Tfdp2	21781/ 211586	E2F1/DP1, G1/S transition	QQIAFKNL*	7.9	0.004
	Pin1	23988	Regulation of cell proliferation	RVYYFNHI	3.0	0.02
Chromatin related	*Ash1l*	192195	Histone methyltransferase	VGLINKDSV	3.6	0.04
	Bptf	207165	Chromatin-mediated regulation of transcription	KAVDFDGRL	4.7	0.004
	Rfc2	19718	PCNA loader, DNA replication	AVLRYTKL	5.3	0.02
	Supt16h	114741	Chromatin-mediated regulation of transcription	VSYKNPSLM*	8.9	0.01
Transcription	*Myb*	17863	Regulates proliferation of hematopoietic cells	NAIKNHWNSTM	4.3	0.01
	Taf6	21343	Subunit of TIIFD	AQQVNRTTL	4.1	0.003
	Znrd1	66136	Component of the RNA polymerase I (Pol I)	TSVVFNKL	2.9	0.01
Signal transduction	*4921505C17Rik*	78757	Component of mTORC2	KALSYASL*	9.7	0.003
	Erbb2ip	59079	Binding partner of Erbb2	GSLKNVTTL	4.6	0.03
Apoptosis	*Sgk1*	20393	Response to DNA damage stimulus	STLTYSRM	9.2	0.03
	Tnfsf10	22035	TNF-related apoptosis inducing ligand (TRAIL)	VSVTNEHLM*	9.0	0.007

Gene ID and Gene Symbol description refer to NCBI gene entries. Fold change and P-values were calculed from biological replicate experiments (n=3). Rapa = rapamycin; CTRL = control. Functional classification is based upon bibliographic searches. *Indicates peptides only detected 48 hours following rapamycin exposure.

We next sought to determine whether differential expression of MHC I peptides following mTOR inhibition correlates with changes in abundance of source proteins. We selected peptide candidates derived from well-characterized source proteins with biological functions closely related to the mTOR signaling pathway. Expression levels of source proteins following mTOR inhibition were quantified using Western blot analysis or flow cytometry. These kinetic profiles were also compared to those observed for their corresponding MHC I peptides (Figure 5.7B and C). For example, the replication factor subunit 2 (Rfc2) and myeloblastosis oncogene (Myb) showed a 1.8-fold and 11-fold increase in abundance following treatment with rapamycin, respectively. Both proteins are involved in processes affected by mTOR inhibition; Myb is an oncogene that regulates proliferation of hematopoietic cells and Rfc2 is a protein implicated in DNA replication [30, 31]. The cell surface expression of protein tumor necrosis factor superfamily member 10 (Tnfsf10) also known as TRAIL showed a 1.3-fold decrease over the same period. Interestingly, we observed a 9-fold increase in the expression of MHC 1 peptide from Tnfsf10 between 24 and 48 hours correlating with the decrease in abundance of its corresponding source protein. Previous investigations indicated that rapamycin blocked TRAIL gene expression, which inhibits p70 S6 kinase involved in CD28 and interleukin (IL)-2 receptor signaling [32]. Tnfsf10 plays a role in suppressing tumor initiation and is currently being developed as a specific chemotherapeutic in transformed or stressed cells [33]. In contrast, the proteins menage-a-trois 1 (Mnat1) and rapamycin-insensitive companion of mTOR (RICTOR) showed no significant change in abundance although we observed a marked increased in their corresponding MHC I peptides. These results suggest that no direct correlation could be made between steady-state level of source protein and peptide expression, supporting the notion that only a weak relationship exists between the proteome and the MHC I peptide abundance at the cell surface [34].

5.4 Discussion

We developed a high-throughput MS-based quantitative method to monitor the interplay between the MIP repertoire and the mTOR signaling pathway. Our study reports for the first time the impact of the mTOR oncogenic pathway on the genesis and the composition of the MIP repertoire of neoplastic thymocytes. We found that ~70% of the MIP repertoire is overexpressed following prolonged mTOR inhibition. These results indicates that even if mTOR inhibition is known to decrease protein synthesis [14], other important biological mechanisms taking place in neoplastic cells can lead to an increase of the MHC I peptide pool presented at the cell surface. FACS analyses performed on EL-4 cells exposed to rapamycin for 48 h indicated an overall 30 to 60% increase in cell surface expression of MHC I molecules. While the higher expression levels of MHC I molecules accounts in part for the increase in immunopeptidome population, other mechanisms are likely to participate in the peptide presentation following mTOR inhibition. Indeed, we found that tapasin is overexpressed in EL4 cells following prolonged mTOR inhibition (Figure 5.5B). This chaperone protein plays an important role in the selection of peptides binding to MHC I molecules [6] and is down-regulated in many type of cancers [29]. Our data suggest that EL4 neoplastic cells normally express low levels of tapasin, accounting in part for the reduction in peptide presentation and decreased immune response. Inhibition of the mTOR signaling pathway by rapamycin may contributes to restoring tapasin expression to levels comparable to those of normal thymocytes. Furthermore, we also observed an increase in tapasin levels in other neoplastic cells (e.g. RMA) exposed to rapamycin (data not shown). Because tapasin overexpression has been shown to be sufficient to induce upregulation of MHC I expression [29], our results led us to propose that rapamycin treatment might enhance the immunogenicity of neoplastic cells. However, further studies on diverse cancer types expressing low levels of tapasin are needed to further understand the effect of rapamycin on neoplastic cells.

The broader effect of rapamycin on the mTOR signaling pathway could also influence peptide degradation and processing by favoring autophagy, a cellular pathway essential for survival, differentiation, development, and homeostatis [35, 36]. Recent reports suggest that diverse cellular mechanisms, including autophagy, participate in

antigen presentation [1]. We found that autophagy was up-regulated in EL4 cells following prolonged mTOR inhibition (Figure 5.1A). These observations suggest that autophagy may affect MHC I peptide presentation by enhancing degradation of proteins derived from mTOR inhibition, and consequently increasing the peptide pool available for MHC I complex assembly. Finally, we suggest that rapamycin treatment might increase the pool of defective ribosomal products (DRiPs) [12]. Yewdell et *al.* have proposed that peptides presented by MHC I molecules originate mainly from DRiPs, which are defective forms of gene products more rapidly degraded than functional products [37]. Consequently, increasing the pool of DRiPs may potentially enhance antigen presentation.

Genomic and proteomic changes taking place within the cell mainly control tumor progression. In many type of cancers, components that control cell cycle progression and protein translation are deregulated, leading to an inappropriate cellular proliferation [38]. We found that the MIP repertoire of EL4 neoplastic thymocytes was enriched in peptides derived from proteins implicated in biological processes such as regulation of cell cycle, ribosome biogenesis, and translation (Figure 5.3). Our data suggest that the MIP repertoire of EL4 cells may conceals an imprint of its neoplastic origin tissue. EL4 cells represent a meaningful model to study the effect of the mTOR pathway on the MIP repertoire. The mTOR pathway plays an essential role in cancer cell growth and proliferation [16]. We found that ~55% of differentially expressed MHC I peptides following mTOR inhibition are derived from genes involved in cell cycle progression and/or cell proliferation (Table 5.1, Table AIII.III). Interestingly, overexpression of some MHC I peptides derived from specific genes such as *Mnat1*, *Tfdp1/2*, *Erbin*, and *Pin1* can be related to the G1 arrest resulting from rapamycin treatment [19, 39]. In addition, mTOR inhibition also lead to the expression of unique MHC I peptides not typically presented in high abundance at the cell surface of EL4 cells. For example, we found that RICTOR-derived peptide was overexpressed ~10-fold on rapamycin-treated cells (Table I). RICTOR is an important component of the mTOR complex 2 (mTORC2), which is insensitive to rapamycin inhibition. However, a recent report revealed that prolonged rapamycin treatment might decrease mTORC2 by competing for newly synthesized mTOR [40]. This may explain why the RICTOR-derived peptide is only expressed following prolonged mTOR inhibition. Taken together, our immunopeptidome analyses led us to suggest that metabolic processes

taking place following mTOR inhibition imprint the MIP repertoire of neoplastic thymocytes.

Evidences suggest that a high proportion of MHC I peptides are derived from newly synthesized but rapidly degraded proteins (RDPs), also refered to as defective ribosomal products (DRiPs) [2, 12, 22, 41]. Generally, we found that changes in abundance of source proteins following mTOR inhibition do not correlate with differential expression of their corresponding MHC I-associated peptides (Figure 5.7). Our data favor the model that newly synthesized proteins and/or protein degradation rates could affect more significantly peptide generation following mTOR inhibition. Experiments are currently underway to isolate translationally active polysomal complexes [42, 43] from EL4 rapamycin-treated cells to further understand the impact of protein translation on the composition of the MIP repertoire. We hypothesize that this analysis will provide a more representative view of mRNAs that are differentially translated following mTOR inhibition, thereby providing a better correlation between expression of actively translated mRNA and peptide abundance.

5.5 Materials and methods

5.5.1 Cell Lines, Cell culture and Flow Cytometry

The EL4 (β_2m^+ wildtype) and C4.4-25$^-$ (β_2m^- mutant) cell lines were kindly provided by Dr. Rickard Glas (Karolinska Institutet, Karolinska University Hospital, Huddinge, Sweden). These cell lines were grown at 37°C with 5% CO_2 in medium consisting of 75% X-Vivo15/25% DMEM supplemented with 2.5% FBS and antibiotics. EL4 wild type cells were treated with 20 ng/mL of rapamycin (Sigma-Aldrich) for 48, 24, 12 and 6 hours in a reversed time course (T48, T24, T12, T6, T0) to obtain one equivalent control (T0) for each condition. MHC I peptide and protein extractions were conducted at T0 for the three different cultures (biological replicates). MHC class I molecules at the cell surface were stained anti-H-2K^b (Y3, ATCC), anti-H-2D^b (clone B22-249.R1, CedarLane), and biotin-conjugated anti-Qa-2 (clone 1-1-2, BD Pharmingen) and analyzed on a BD LSR II flow cytometer using FACSDiva software (BD Biosciences). Estimation of apoptotic cells and TRAIL positive cells were based on staining with PE-conjugated annexin V (BD Biosciences) and FITC-conjugated TRAIL antibody (eBioscience), respectively. Cell number and cell size was measured by light microscopy using the Axio Imager (Zeiss) and the Northern Eclipse software (Empix Imaging Inc.).

5.5.2 Western Blot Analysis

After rapamycin treatment, the EL4 cells (1-3 X 10^7 cells/condition) were lysed with 400 µL of Triton buffer containing 0.5% Triton, 50mM Tris-HCl (pH 7.5), 5 mM EDTA, 200mM NaCl, 1 mM sodium orthovanadate, 1 mM sodium molybdate, 2 mM imidazole, 4 mM sodium tartrate dihydrate and protease inhibitors (1:100, Sigma). Cells were incubated on ice for 30 mins and centrifuged at 10,000g to pellet cellular debris. Protein amount determination was performed using the BCA assay (Pierce). Protein samples (20 µg of total lysate/lane) were resolved by 1D gel electrophoresis using NUPAGE 4-12% Bis-Tris midi gels (Invitrogen) according to the manufacturer's instructions and blotted on nitrocellulose membranes. Solution of TBS containing 0.1% Tween 20 (TBST) and 5% (w/v) nonfat dry milk was used to block non-specific binding

with nitrocellulose. The membranes were incubated with primary antibodies including Tapasin, Lmp7, Lmp2, ERp57, β1, β5 (Abcam); Calnexin, β-actin (Sigma); PA28, p70 S6 kinase (p70S6K), phospho-p70S6K (Ser^{424}/Thr^{421}), phospho-p70S6K (Thr^{389}), 4E-BP1, phospho-4E-BP1 (Ser^{65}), α-tubulin, RICTOR (Cell Signaling); Myb (Upstate, Millipore); Tap1, Tap2, Mnat1 (Santa Cruz Biotechnology); Mecl-1, β2 (Biomol International); Rfc2 (Proteintech Group, PTG); LC3 (Affinity Bioreagents, ABR). Secondary antibodies used were anti-rabbit HRP and anti-mouse HRP (1:5000, Chemicon International) and anti-goat HRP (1:5000, BD Pharmingen). Blots were probed overnight at 4°C with primary antibodies in TBST (5% milk), washed with TBST, probed with their respective secondary antibodies in TBST (5% milk), washed again with TBST and developed by using enhanced chemiluminescence (ECL, ECL plus, GE Healthcare). For normalization, blots were treated with stripping buffer (Chemicon International) for 20 mins at RT and reprobed with antibody against parent protein, α-tubulin or β-actin. Western Blots showing significant changes were done at least in duplicate from different biological extracts. Development of Western blot was performed using LAS-3000 (Fujifilm) and band intensities were measured and normalized against α-tubulin or β-actin using Multi-Gauge (Fujifilm).

5.5.3 Peptide Extraction and Mass Spectrometry Analysis

Cell surface MHC I peptides were isolated from viable cells as described previously using mild acid elution [10]. The kinetic on EL4 cells were conducted in three biological replicates. Equivalent amount of MHC I material (5 X 10^8 cells/sample) were fractionated using SCX (Strong Cation Exchange) chromatography using an Agilent 1100 capillary system. Sample were loaded onto a home made SCX column (300 μm i.d. X 50 mm, PolyLC material) at 8 uL/min and fractionated into five fractions using a linear gradient from 0-25% buffer B in 25 mins [buffer A: 5 mM ammonium formate (Sigma-Aldrich), 15% acetonitrile (ACN, Fisher Scientific), pH 3.0; buffer B: 2M ammonium formate, 15% ACN, pH 3.0]. Fractions were dried under vacuum and analyzed using nanoLC-MS methods as described previously [10]. Briefly, peptides were loaded on a C_{12} precolumn (Jupiter Proteo 4 μm, Phenomenex) and separated on a homemade reversed phase column (150 μm i.d. X 100 mm, Jupiter Proteo 4 μm particle) at a flow rate of 600 nL/min using a linear gradient of 3-60% aqueous ACN (0.2% formic acid, EM Science) in 69 mins. An

Eksigent LC system was coupled to a LTQ-Orbitrap mass spectrometer (Thermo Electron) and MS and tandem MS experiments were conducted as previously described [10]. The "lock mass" option was selected using the siloxane ion, m/z = 445.120560.

5.5.4 Peptide Detection and Clustering

Raw data files generated from the Orbitrap were processed using in-house peptide detection software Mass Sense to identify all ions according to their specific coordinates (m/z, retention time, charge and intensity). Lists of detected peptides ions (peptide maps) were generated for each individual LC-MS analysis using an intensity threshold of 10,000 above the backgroung noise. Identification files generated from Mascot (Matrix Science) were converted in excel format and peptide lists were filtered using an in-house script to select peptides displaying a MHC I motif associated to $H2D^b$, $H2K^b$, Qa2 and Qa1 alleles [10]. Peptides coordinates (theoretical mass and Rt) reported by Mascot for each MHC I peptide candidates were than aligned with peptide maps using in-house clustering software with tolerances fixed to ± 0.02 m/z, ± 1 min and ± 1 fraction for 2D-LC-MS Orbitrap experiments to generate lists of non-redundant peptide clusters.

5.5.5 MS/MS Sequencing and Protein Identification

The data were searched against International Protein Index (IPI) mouse database using Mascot (Matrix Science) search engine with tolerance of ± 0.05 Da on precursor mass and ± 0.5 Da on fragment ions. Searches were performed without enzyme specificity. Inspection of MS/MS spectra were done manually as previously described [10] and all peptide candidates validated had mass accuracy within 10 ppm of the theoretical values. Peptides structures were validated as MHC I peptides using the β_2m^- control samples ($WT/\beta_2m^- > 5$ or only in WT). List of MHC I peptides identified was blasted against non-redundant NCBInr (mouse entries) using an in-house interface to identify proteins according to the Gene ID annotation.

5.6 Acknowledgements

We are grateful to the staff of the following core facilities at the Institute for Research in Immunology and Cancer for their outstanding support: Animal facility, Bioinformatics, Flow cytometry, and Proteomics. We acknowledge Mathieu Courcelles for his assistance in the development of bioinformatic tools and Danielle de Verteuil for providing some Western Blots. This work was supported by funds from the Canadian Cancer Society and the Terry Fox Foundation through the National Cancer Institute of Canada. MHF and EC are supported by training grants from the FRSQ (Fonds de la Recherche en Santé du Québec) and the Canadian Institutes of Health Research, respectively. CP and PT hold Canada Research Chairs in Immunobiology, and Proteomics and Bioanalytical Mass Spectrometry, respectively. The authors have no conflicting financial interests.

5.7 References

1. Vyas, J.M., Van der Veen, A.G., and Ploegh, H.L., *The known unknowns of antigen processing and presentation.* Nat Rev Immunol, 2008. **8**(8): p. 607-18.

2. Yewdell, J.W., Reits, E., and Neefjes, J., *Making sense of mass destruction: quantitating MHC class I antigen presentation.* Nat Rev Immunol, 2003. **3**(12): p. 952-61.

3. Cresswell, P., Ackerman, A.L., Giodini, A., Peaper, D.R., and Wearsch, P.A., *Mechanisms of MHC class I-restricted antigen processing and cross-presentation.* Immunol Rev, 2005. **207**: p. 145-57.

4. Elliott, T. and Williams, A., *The optimization of peptide cargo bound to MHC class I molecules by the peptide-loading complex.* Immunol Rev, 2005. **207**: p. 89-99.

5. Hammer, G.E., Kanaseki, T., and Shastri, N., *The final touches make perfect the peptide-MHC class I repertoire.* Immunity, 2007. **26**(4): p. 397-406.

6. Hammer, G.E. and Shastri, N., *Construction and destruction of MHC class I in the peptide-loading complex.* Nat Immunol, 2007. **8**(8): p. 793-4.

7. Anikeeva, N., Lebedeva, T., Clapp, A.R., Goldman, E.R., Dustin, M.L., Mattoussi, H., and Sykulev, Y., *Quantum dot/peptide-MHC biosensors reveal strong CD8-dependent cooperation between self and viral antigens that augment the T cell response.* Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16846-51.

8. Dunn, G.P., Old, L.J., and Schreiber, R.D., *The immunobiology of cancer immunosurveillance and immunoediting.* Immunity, 2004. **21**(2): p. 137-48.

9. Reits, E.A., Hodge, J.W., Herberts, C.A., Groothuis, T.A., Chakraborty, M., Wansley, E.K., Camphausen, K., Luiten, R.M., de Ru, A.H., Neijssen, J., Griekspoor, A., Mesman, E., Verreck, F.A., Spits, H., Schlom, J., van Veelen, P., and Neefjes, J.J., *Radiation modulates the peptide repertoire, enhances MHC class I expression, and induces successful antitumor immunotherapy.* J Exp Med, 2006. **203**(5): p. 1259-71.

10. Fortier, M.H., Caron, E., Hardy, M.P., Voisin, G., Lemieux, S., Perreault, C., and Thibault, P., *The MHC class I peptide repertoire is molded by the transcriptome.* J Exp Med, 2008. **205**(3): p. 595-610.

11. Yewdell, J.W., *Plumbing the sources of endogenous MHC class I peptide ligands.* Curr Opin Immunol, 2007. **19**(1): p. 79-86.

12. Yewdell, J.W. and Nicchitta, C.V., *The DRiP hypothesis decennial: support, controversy, refinement and extension.* Trends Immunol, 2006. **27**(8): p. 368-73.

13. Chiang, G.G. and Abraham, R.T., *Targeting the mTOR signaling network in cancer.* Trends Mol Med, 2007. **13**(10): p. 433-42.

14. Guertin, D.A. and Sabatini, D.M., *Defining the role of mTOR in cancer.* Cancer Cell, 2007. **12**(1): p. 9-22.

15. Levine, B. and Kroemer, G., *Autophagy in the pathogenesis of disease.* Cell, 2008. **132**(1): p. 27-42.

16. Yang, Q. and Guan, K.L., *Expanding mTOR signaling.* Cell Res, 2007. **17**(8): p. 666-81.

17. Yap, T.A., Garrett, M.D., Walton, M.I., Raynaud, F., de Bono, J.S., and Workman, P., *Targeting the PI3K-AKT-mTOR pathway: progress, pitfalls, and promises.* Curr Opin Pharmacol, 2008. **8**(4): p. 393-412.

18. Paglin, S., Lee, N.Y., Nakar, C., Fitzgerald, M., Plotkin, J., Deuel, B., Hackett, N., McMahill, M., Sphicas, E., Lampen, N., and Yahalom, J., *Rapamycin-sensitive pathway regulates mitochondrial membrane potential, autophagy, and survival in irradiated MCF-7 cells.* Cancer Res, 2005. **65**(23): p. 11061-70.

19. Easton, J.B. and Houghton, P.J., *mTOR and cancer therapy.* Oncogene, 2006. **25**(48): p. 6436-46.

20. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

21. Hickman-Miller, H.D. and Hildebrand, W.H., *The immune response under stress: the role of HSP-derived peptides.* Trends Immunol, 2004. **25**(8): p. 427-33.

22. Schubert, U., Anton, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., and Bennink, J.R., *Rapid degradation of a large fraction of newly synthesized proteins by proteasomes.* Nature, 2000. **404**(6779): p. 770-4.

23. Tabaczewski, P., Shirwan, H., Lewis, K., and Stroynowski, I., *Alternative splicing of class Ib major histocompatibility complex transcripts in vivo leads to the expression of soluble Qa-2 molecules in murine blood.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1883-7.

24. Kloetzel, P.M., *Antigen processing by the proteasome.* Nat Rev Mol Cell Biol, 2001. **2**(3): p. 179-87.

25. Grolleau, A., Bowman, J., Pradet-Balade, B., Puravs, E., Hanash, S., Garcia-Sanz, J.A., and Beretta, L., *Global and specific translational control by rapamycin in T cells uncovered by microarrays and proteomics.* J Biol Chem, 2002. **277**(25): p. 22175-84.

26. Peng, T., Golub, T.R., and Sabatini, D.M., *The immunosuppressant rapamycin mimics a starvation-like signal distinct from amino acid and glucose deprivation.* Mol Cell Biol, 2002. **22**(15): p. 5575-84.

27. Jensen, P.E., *Recent advances in antigen processing and presentation.* Nat Immunol, 2007. **8**(10): p. 1041-8.

28. Sadegh-Nasseri, S., Chen, M., Narayan, K., and Bouvier, M., *The convergent roles of tapasin and HLA-DM in antigen presentation.* Trends Immunol, 2008. **29**(3): p. 141-7.

29. Lou, Y., Basha, G., Seipp, R.P., Cai, B., Chen, S.S., Moise, A.R., Jeffries, A.P., Gopaul, R.S., Vitalis, T.Z., and Jefferies, W.A., *Combining the antigen processing components TAP and Tapasin elicits enhanced tumor-free survival.* Clin Cancer Res, 2008. **14**(5): p. 1494-501.

30. Johnson, A., Yao, N.Y., Bowman, G.D., Kuriyan, J., and O'Donnell, M., *The replication factor C clamp loader requires arginine finger sensors to drive DNA binding and proliferating cell nuclear antigen loading.* J Biol Chem, 2006. **281**(46): p. 35531-43.

31. Lieu, Y.K., Kumar, A., Pajerowski, A.G., Rogers, T.J., and Reddy, E.P., *Requirement of c-myb in T cell development and in mature T cell function.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14853-8.

32. Musgrave, B.L., Phu, T., Butler, J.J., Makrigiannis, A.P., and Hoskin, D.W., *Murine TRAIL (TNF-related apoptosis inducing ligand) expression induced by T cell activation is blocked by rapamycin, cyclosporin Á, and inhibitors of phosphatidylinositol 3-kinase, protein kinase C, and protein tyrosine kinases: evidence for TRAIL induction via the T cell receptor signaling pathway.* Exp Cell Res, 1999. **252**(1): p. 96-103.

33. Cretney, E., Shanker, A., Yagita, H., Smyth, M.J., and Sayers, T.J., *TNF-related apoptosis-inducing ligand as a therapeutic agent in autoimmunity and cancer.* Immunol Cell Biol, 2006. **84**(1): p. 87-98.

34. Milner, E., Barnea, E., Beer, I., and Admon, A., *The turnover kinetics of major histocompatibility complex peptides of human cancer cells.* Mol Cell Proteomics, 2006. **5**(2): p. 357-65.

35. Baehrecke, E.H., *Autophagy: dual roles in life and death?* Nat Rev Mol Cell Biol, 2005. **6**(6): p. 505-10.

36. Mathew, R., Karantza-Wadsworth, V., and White, E., *Role of autophagy in cancer.* Nat Rev Cancer, 2007. **7**(12): p. 961-7.

37. Yewdell, J.W., Schubert, U., and Bennink, J.R., *At the crossroads of cell biology and immunology: DRiPs and other sources of peptide ligands for MHC class I molecules.* J Cell Sci, 2001. **114**(Pt 5): p. 845-51.

38. Bartek, J. and Lukas, J., *Cell cycle. Order from destruction.* Science, 2001. **294**(5540): p. 66-7.

39. Fingar, D.C. and Blenis, J., *Target of rapamycin (TOR): an integrator of nutrient and growth factor signals and coordinator of cell growth and cell cycle progression.* Oncogene, 2004. **23**(18): p. 3151-71.

40. Sarbassov, D.D., Ali, S.M., Sengupta, S., Sheen, J.H., Hsu, P.P., Bagley, A.F., Markhard, A.L., and Sabatini, D.M., *Prolonged rapamycin treatment inhibits mTORC2 assembly and Akt/PKB.* Mol Cell, 2006. **22**(2): p. 159-68.

41. Reits, E.A., Vos, J.C., Gromme, M., and Neefjes, J., *The major substrates for TAP in vivo are derived from newly synthesized proteins.* Nature, 2000. **404**(6779): p. 774-8.

42. Bilanges, B., Argonza-Barrett, R., Kolesnichenko, M., Skinner, C., Nair, M., Chen, M., and Stokoe, D., *Tuberous sclerosis complex proteins 1 and 2 control serum-dependent translation in a TOP-dependent and -independent manner.* Mol Cell Biol, 2007. **27**(16): p. 5746-64.

43. Preiss, T., Baron-Benhamou, J., Ansorge, W., and Hentze, M.W., *Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock.* Nat Struct Biol, 2003. **10**(12): p. 1039-47.

6. Conclusion

6.1 Conclusion générale

Les peptides présentés par les molécules MHC I jouent un rôle important au niveau de la régulation du développement des cellules T et de l'immunosurveillance. Certains peptides qui sont présentés majoritairement sur les cellules leucémiques peuvent être utilisés comme cibles potentielles pour le traitement de la leucémie par l'immunothérapie. Cependant, le ciblage spécifique de ces peptides est jusqu'à présent limité par nos connaissances fragmentaires concernant la diversité de la population allogénique présentée à la surface des cellules leucémiques. L'objectif principal de cette thèse était donc de développer une nouvelle méthodologie analytique pour fournir une définition moléculaire précise du répertoire MIP présenté par des cellules normales et leucémiques.

Étant donné que les peptides MHC I sont présents en quantité infime à la surface des cellules, l'évaluation de différents systèmes chromatographiques fut effectuée dans un premier temps pour tenter d'augmenter la sensibilité et la sélectivité de la séparation chromatographique (chapitre 3). Dans cette optique, les performances analytiques des colonnes capillaires (nanoLC-MS) furent comparées à celles d'un nouveau système microfluidique (nanoLC-chip-MS, Agilent Technologies). Le concept modulaire de la puce nanoLC permet de réduire les volumes morts par la diminution des lignes de transfert et des connexions se traduisant par une amélioration des performances chromatographiques. Les avantages et les limitations des deux systèmes en terme de capacité chromatographique, de reproductibilité et de sensibilité furent comparés à partir d'analyses effectuées sur des mélanges protéiques complexes incluant des échantillons plasmiques. L'analyse des profils peptidiques à travers différentes injections a révélé que la puce nanoLC permettait d'obtenir d'excellentes reproductibilités au niveau des temps de rétention et des intensités avec des valeurs de RSD inférieures à 0,5 et 9,1%, respectivement. Les expériences ont démontré que des quantités de l'ordre du 1 à 5 fmol de protéines digérées pouvaient être détectées au sein de matrices plasmiques complexes. De plus, des variations d'abondance représentant un facteur 2 ont pu être détectées pour des quantités très faibles de protéines digérées (2 à 5 fmol). La configuration du système microfluidique a facilité aussi l'intégration de la chromatographie bidimensionnelle permettant d'augmenter la sélectivité et la capacité de chargement, facilitant ainsi la détection de protéines présentées en faible abondance (e.g.

insulin-like growth factor binding protein 1). Ces études ont permis d'établir que les performances des deux systèmes étudiés sont comparables en terme de limite de détection ainsi qu'au niveau de la reproductibilité des temps de rétention et des intensités. Étant donné que l'instrumentation MS pouvant être couplée à la puce nanoLC était plutôt limitée au moment de l'optimisation de la méthode analytique pour l'étude du répertoire MIP, l'approche utilisant les colonnes capillaires fut préférée pour la continuation du projet. Ce système fut par contre employé dans le cadre d'un autre projet pour l'analyse différentielle du phosphoprotéome de cellules U937 suite à l'exposition au PMA [1].

Dans le cadre des études présentées dans le chapitre 4, nous avons proposé une nouvelle méthodologie analytique combinant différentes stratégies bioinformatiques aux techniques analytiques de séparation par 2D-nanoLC-MS pour résoudre la complexité associée au répertoire MIP. La combinaison de cette stratégie MS avec l'utilisation de l'élution acide pour l'extraction des peptides MHC I a permis la génération d'un portrait détaillé du répertoire MIP présenté par $< 10^8$ de cellules thymiques EL4. Cette méthodologie analytique permet donc l'identification d'un nombre important de peptides MHC I à partir d'une quantité limitée de cellules. L'approche de quantification peut d'ailleurs être appliquée à l'étude du répertoire MIP sur tous les modèles cellulaires. Les résultats ont indiqué que l'utilisation d'un contrôle β_2m^- permettait une discrimination efficace entre les structures peptidiques associées au MHC I et celles dérivant des peptides contaminants qui sont présentées au sein des mélanges complexes obtenus de l'élution acide. Les travaux ont montré que l'extraction par élution acide pouvait aussi être utilisée pour l'étude du répertoire MIP présenté par des cellules primaires. L'utilisation de critères stricts de sélection par l'intermédiaire des algorithmes de prédiction a permis d'obtenir des taux de fausses identifications ~2%.

Cette étude menée chez la souris a conduit à l'identification respective de 189 et 196 peptides MHC I présentés par des cellules normales et néoplasiques de thymocytes. L'intégration des données peptidiques avec celle du transcriptome du thymus a mené à la conclusion que le répertoire MIP de cellules normales thymiques contenait une signature spécifique de son tissus d'origine étant associée à ~17% des gènes représentés dans le répertoire MIP. De plus, les résultats ont mis en évidence que le répertoire MIP des

thymocytes était enrichi en peptides dérivant de transcrits très abondants et en protéines sources provenant de la famille des cyclines et des hélicases. La comparaison des répertoires MIP des cellules normales et leucémiques a permis de constater des variations d'expression significatives pour environ 25% de la population peptidique MHC I. De ce nombre, plus de la moitié des peptides provenaient de protéines sources directement impliquées au niveau de la transformation néoplasique (e.g. composantes de la voie de signalisation PI3K-AKT-mTOR). La comparaison des rapports d'expression entre les cellules normales et néoplasiques de thymocytes pour les peptides et les transcrits sources de peptides a permis de constater une corrélation modeste ($r = 0.63$) entre les changements observés au niveau de l'expression des peptides et des transcrits. Une plus faible corrélation ($r = 0.32$) avait préalablement été déterminée par Weinzierl et *al.* [2]. Cette différence peut probablement s'expliquer par le fait que l'estimation du niveau de transcrits fut effectuée par qRT-PCR dans notre étude plutôt que par l'utilisation des microarrays. Cependant, les résultats que nous avons obtenus supportent la principale conclusion proposée par Weinzierl et *al.* qui stipule que la majorité des peptides surexprimés sur les cellules néoplasiques (74% dans notre étude) n'auraient pas été identifiés par l'estimation du niveau de transcrits. Ces résultats suggèrent que la surexpression des peptides MHC I présentés par les cellules néoplasiques est causée par des mécanismes post-traductionnels, et que ces candidats peuvent seulement être détectés à l'aide d'approches immunoprotéomiques quantitatives. Les essais d'immunisation effectués sur les peptides VAAANREVL et STLTYSRM, qui sont surexprimés sur les cellules néoplasiques, ont d'ailleurs mené à la conclusion que des changements en expression de l'ordre de 10 à $\geq$ 85 fois pouvaient stimuler une réponse cytotoxique spécifique. Ces résultats supportent le concept que l'analyse à haut débit par spectrométrie de masse du répertoire MIP est une plateforme des plus intéressantes pour l'identification d'antigènes tumoraux immunogéniques.

Dans le cadre des travaux présentés dans le chapitre 5, la méthodologie analytique fut aussi employée pour étudier la cinétique de présentation MHC I suite au traitement à la rapamycine, molécule étant connue pour inhiber le mécanisme TOR qui a une influence majeure dans le processus oncogénique. La voie de signalisation mTOR est connue pour jouer un rôle déterminant dans la régulation de la synthèse et de la dégradation des protéines [3], deux processus dynamiques que les évidences suggèrent comme ayant une influence majeure sur la composition du répertoire MIP [4]. Ces études ont mené à

l'identification de 593 peptides MHC I présentés à la surface de cellules néoplasiques de thymocytes (EL4). La comparaison des populations peptidiques MHC I à travers les différents temps d'incubation a permis d'observer une augmentation en expression pour 70% du répertoire MIP après une exposition prolongée à la rapamycine (48 heures). Ces résultats sont en concordance avec une augmentation de l'expression de surface des molécules MHC I, suggérant un changement au niveau de la génération ou de la présentation antigénique en présence du stress causé par la rapamycine. De ce nombre, 17% des peptides ont été identifiés comme étant significativement surexprimés suite à une exposition prolongée à la rapamycine, d'où plus de 50% des protéines sources sont impliquées dans la prolifération et au niveau de la progression du cycle cellulaire. La comparaison entre les profils d'abondance de peptides MHC I surexprimés et de leurs protéines sources associées a mené à la conclusion que le répertoire MIP ne reflète pas le protéome de la cellule. Ces résultats suggèrent plutôt que des changements au niveau des taux de synthèse *de novo* des protéines et/ou de la dégradation des protéines affecteraient de façon plus importante la composition du répertoire MIP suite à l'inhibition de mTOR.

Différentes hypothèses furent proposées pour expliquer l'augmentation générale de la présentation antigénique suite à l'exposition prolongée à la rapamycine. L'analyse des profils d'expression pour les différentes protéines impliquées dans la présentation MHC I a révélé que la chaperonne Tapasin était surexprimée suite au traitement à la rapamycine. Cette protéine est souvent faiblement exprimée dans différents types de cancers ce qui a pour conséquence de diminuer l'expression MHC I [5]. Notre première hypothèse propose donc que la rapamycine augmente le niveau de Tapasin dans les cellules EL4, ce qui restaure l'expression MHC I à la surface. Comme deuxième hypothèse, nous proposons que l'autophagie joue potentiellement un rôle dans la présentation MHC I [6] en augmentant la dégradation des protéines endommagées ou superflues en lien avec l'inhibition de mTOR, ce qui aurait pour conséquence d'augmenter la quantité de peptides disponibles pour l'assemblage des molécules MHC I. Finalement, nous suggérons comme dernière hypothèse que le traitement à la rapamycine régulerait positivement la quantité de DRiPs [7] dans les cellules EL4 ce qui mènerait à une augmentation de la présentation antigénique. Différentes expériences sont d'ailleurs en cours présentement pour valider ces hypothèses (discuté dans la section suivante). D'un point de vue pratique, la régulation positive de la présentation antigénique par la rapamycine pourrait être exploitée pour

augmenter le caractère immunogénique de certains peptides cibles pouvant être utilisés pour l'immunothérapie.

Dans l'ensemble, les résultats obtenus dans le cadre de ces travaux de recherche mettent en valeur l'importance de l'information unique générée par les analyses à haut débit MS, qui résulte en une meilleure compréhension des mécanismes qui régulent la composition et de la genèse du répertoire MIP de cellules normales et néoplasiques.

6.2 Travaux futurs

Différentes expériences biologiques sont présentement en cours dans l'objectif de mieux comprendre les mécanismes qui influencent la présentation antigénique suite à l'inhibition de mTOR par la rapamycine (projet chapitre 5). Des résultats préliminaires obtenus pour la Tapasin ont révélé que la rapamycine augmente l'expression de la protéine chaperonne dans d'autres lignées cancéreuses. Des expériences sont présentement en cours pour transfecter les cellules EL4 avec la Tapasin et évaluer si on observe une corrélation entre le niveau d'expression de la protéine et la présentation MHC I. D'autres analyses sont effectuées en parallèle pour quantifier biochimiquement la quantité de DRiPs présentée dans les cellules EL4 suite à l'exposition prolongée à la rapamycine [8].

Pour mieux comprendre l'impact de traduction des protéines sur la composition du répertoire MIP, des expériences sont actuellement en cours pour isoler les polysomes des cellules EL4 suite à l'inhibition de mTOR [9, 10]. Nous émettons l'hypothèse qu'une meilleure corrélation linéaire pourrait être observée entre l'abondance des peptides et l'expression des transcrits actifs (polysomes). Nous suggérons que cette expérience pourrait améliorer notre niveau de connaissance par rapport à la genèse du répertoire MIP.

La méthodologie analytique développée dans cette thèse est actuellement utilisée dans le cadre d'un autre projet consistant à évaluer l'influence de l'immunoprotéasome [11] sur la composition du répertoire MIP présentée à la surface de cellules dendritiques. Cette expérience permettra une meilleure compréhension du rôle de l'immunoprotéasome dans la génération des peptides antigéniques. L'approche analytique sera aussi employée pour étudier le répertoire MIP de cellules cancéreuses chez l'humain.

6.3 Références

1. Ghitun, M., Bonneil, E., Fortier, M.H., Yin, H., Killeen, K., and Thibault, P., *Integrated microfluidic devices with enhanced separation performance: application to phosphoproteome analyses of differentiated cell model systems.* J Sep Sci, 2006. **29**(11): p. 1539-49.

2. Weinzierl, A.O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.G., and Stevanovic, S., *Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface.* Mol Cell Proteomics, 2007. **6**(1): p. 102-13.

3. Guertin, D.A. and Sabatini, D.M., *Defining the role of mTOR in cancer.* Cancer Cell, 2007. **12**(1): p. 9-22.

4. Yewdell, J.W., *Plumbing the sources of endogenous MHC class I peptide ligands.* Curr Opin Immunol, 2007. **19**(1): p. 79-86.

5. Lou, Y., Basha, G., Seipp, R.P., Cai, B., Chen, S.S., Moise, A.R., Jeffries, A.P., Gopaul, R.S., Vitalis, T.Z., and Jefferies, W.A., *Combining the antigen processing components TAP and Tapasin elicits enhanced tumor-free survival.* Clin Cancer Res, 2008. **14**(5): p. 1494-501.

6. Vyas, J.M., Van der Veen, A.G., and Ploegh, H.L., *The known unknowns of antigen processing and presentation.* Nat Rev Immunol, 2008. **8**(8): p. 607-18.

7. Yewdell, J.W. and Nicchitta, C.V., *The DRiP hypothesis decennial: support, controversy, refinement and extension.* Trends Immunol, 2006. **27**(8): p. 368-73.

8. Qian, S.B., Bennink, J.R., and Yewdell, J.W., *Quantitating defective ribosome products.* Methods Mol Biol, 2005. **301**: p. 271-81.

9. Bilanges, B., Argonza-Barrett, R., Kolesnichenko, M., Skinner, C., Nair, M., Chen, M., and Stokoe, D., *Tuberous sclerosis complex proteins 1 and 2 control serum-dependent translation in a TOP-dependent and -independent manner.* Mol Cell Biol, 2007. **27**(16): p. 5746-64.

10. Preiss, T., Baron-Benhamou, J., Ansorge, W., and Hentze, M.W., *Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock.* Nat Struct Biol, 2003. **10**(12): p. 1039-47.

11. Kloetzel, P.M., *Antigen processing by the proteasome.* Nat Rev Mol Cell Biol, 2001. **2**(3): p. 179-87.

Annexe I : Script PHP pour la sélection des peptides MHC I.

PHP script for MHC class I peptide selection

```
function isMHCpeptide($pepseq) {
$peplength = strlen($pepseq);
$poslast = substr($pepseq,.$peplength -1, 1);
if ($peplength >= 8 && $peplength <= 9) {
$pos3 = substr($pepseq, 2, 1);
$pos5 = substr($pepseq, 4, 1);
$pos7 = substr($pepseq, 6, 1);
if (preg_match('/[ILVMF]/', $poslast) && ((preg_match('/[YF]/', $pos3) && $peplength == 8) || preg_match('/[YF]/', $pos5) ||
(preg_match('/[H]/', $pos7) && $peplength == 9) || (preg_match('/[N]/', $pos5) && $peplength == 9))) {
return 1;
} else
if ($peplength == 9 && preg_match('/LLL$/', $pepseq)) {
return 1;
}
} else
if ($peplength >= 10 && $peplength <= 13) {
$pos5 = substr($pepseq, 4, 1);
if (preg_match('/[N]/', $pos5) && preg_match('/[ILVMF]/', $poslast)) {
return 1;
}
}
return 0;
}
```

Annexe II : Informations supplémentaires pour le chapitre 4

(The MHC class I peptide repertoire is molded by the transcriptome)

Table AII.I: Relative abundance of peptides eluted from WT, β_2m^- and β_2m^+ EL4 cell lines.

Gene ID	Gene Symbol	Peptide	WT	Intensities from β_2m^-	β_2m^+	Fold change WT/β_2m^-
68523	*1110019N10Rik*	AALENTHLL	160364	12000	39986	13
66185	*1110037F02Rik*	RQILNADAM	364055	12000	227326	30
67463	*1200014M14Rik*	GALENAKAEI	191356	12000	381782	16
74174	*1700006H03Rik*	NSPANNIVM	68002	12000	1534480	6
100910	*2010209O12Rik*	AQIRNLTVL	212465	12000	106874	18
76559	*2410024A21Rik*	AGPENSSKI	216979	12000	649743	18
75425	*2610036D13Rik*	QTVENVEHL	356744	52221	827933	7
67268	*2900073G15Rik*	SMGKNPTDEYL	4278255	12000	11467376	357
77877	*6030458C11Rik*	KVLDVLHSL	63278	12000	71265	5
338523	*A630082K20Rik*	SSIQNGKYTL	1351220	12000	480198	113
18671	*Abcb1a*	STVRNADVI	3232678	12000	12000	269
11307	*Abcg1*	VNIEFKDL	838637	12000	2348761	70
104112	*Acly*	SIANFTNV	59580	12000	225167	5
66713	*Actr2*	YAGSNFPEHI	5117498	12000	6554584	426
67019	*Actr6*	SGYSFTHI	428791	12000	996812	36
226747	*Ahctf1*	TSVENHEFL	676119	12000	1115879	56
223774	*Alg12*	AAMLNYTHI	121626	12000	450762	10
81702	*Ankrd17*	ASVLNVNHI	7537305	12000	14073365	628
11765	*Ap1g1*	FALVNGNNI	238448	12000	508707	20
71435	*Arhgap21*	TGVTNRDLI	421388	12000	194111	35
11984	*Atp6v0c*	SAMVFSAM	100221	12000	314743	8
328099	*AU021838*	VSYLFSHV	163732	12000	285904	14
66165	*Bccip*	KAPVNTAEL	1741908	12000	998607	145
213895	*Bms1l*	YQNQEIHNL	1272300	12000	1633700	106
12176	*Bnip3*	LSMRNTSVM	133906	12000	371093	11
12181	*Bop1*	EQVALVHRL	306162	12000	1219215	26

107976	*Bre*	ATQVYPKL	1147903	12000	3484401	96
224171	*C330027C09Rik*	AQNDIEHLF	1329747	12000	1373696	111
76740	*C920006C10Rik*	SAIINEDNL	899101	12000	1629693	75
69719	*Cad*	RQAENGMYI	92562	12000	1267863	8
12340	*Capza1*	ISFKFDHL	35372	12000	56253	3
26885	*Casp8ap2*	VTVLNVDHL	285120	12000	299518	24
12449	*Ccnf*	SQAVNKQQI	236020	68580	828654	3
12465	*Cct5*	SMMDVDHQI	1326343	12000	2363592	111
12468	*Cct7*	RAIKNDSVV	930640	12000	1300660	78
12576	*Cdkn1b*	FGPVNHEEL	11267481	12000	17543161	939
229841	*Cenpe*	SALQNAESDRL	1125746	12000	2349902	94
216859	*Centb1*	INQIYEARV	691884	12000	1065038	58
12649	*Chek1*	TGPSNVDKL	6545497	112308	14733333	58
214901	*Chtf18*	QLSDTLHSL	61094	12000	190594	5
80986	*Ckap2*	TLNDLIHNI	136301	12000	345332	11
234594	*Cnot1*	IAFIFNNL	46952	12000	74559	4
12847	*Copa*	SQLPVDHIL	404318	12000	676882	34
12877	*Cpeb1*	SMLQNPLGNVL	185198	12000	528268	15
68975	*Crsp8*	SGLLNQQSL	978436	99069	1634916	10
71745	*Cul2*	VINSFVHV	67372	12000	244370	6
215821	*D10Bwg1379e*	AVLTNQETI	158503	12000	747537	13
52635	*D12Ertd551e*	RQLENGTTL	1477468	12000	1284587	123
218978	*D14Ertd436e*	SQHVNLDQL	322974	12000	690276	27
229473	*D930015E06Rik*	ASLVNSPSYL	160789	12000	1730981	13
208846	*Daam1*	AAKVNMTEL	2278053	12000	1980628	190
68087	*Dcakd*	VIQVFQQL	241295	12000	464584	20
67755	*Ddx47*	KTFLFSATM	70060	12000	289017	6
13209	*Ddx6*	INFDFPKL	207697	12000	58611	17
72121	*Dennd2d*	VIYPFMQGL	507408	28423	12000	18
217207	*Dhx8*	QLYTLGAL	185369	12000	505758	15
94176	*Dock2*	SMVQNRVFL	895081	12000	671603	75
233115	*Dpy19l3*	SVVAFHNL	2131426	12000	1928879	178

76843	*Dtl*	SSPENKNWL	2840050	12000	10831315	237
13424	*Dync1h1*	ASYEFVQRL	2201922	33937	1946008	65
66967	*Edem3*	FMATNPEHL	255530	12000	303884	21
13669	*Eif3s10*	QSIEFSRL	3331791	12000	6269562	278
15569	*Elavl2*	GAVTNVKVI	120705	12000	327586	10
14791	*Emg1*	VSVEYTEKM	133779	12000	214864	11
66366	*Ergic3*	QQLDVEHNL	1458352	12000	807276	122
13877	*Erh*	YQPYNKDWI	108556	12000	29068	9
14050	*Eya3*	TIIIFHSL	465796	12000	216132	39
212377	*F730047E07Rik*	VGVTYRTL	84194	12000	462448	7
70611	*Fbxo33*	SSLSNPIANTM	127288	12000	69463	11
14137	*Fdft1*	ISLEFRNL	523739	12000	130109	44
57778	*Fmnl1*	LQYEFTHL	770529	12000	1208086	64
22379	*Fmnl3*	LQYEFTKL	642004	12000	934003	54
108655	*Foxp1*	QQLQQQHLL	705266	12000	1054408	59
239554	*Foxred2*	LGVTNFVHM	168986	12000	118096	14
56784	*Garnl1*	GAIVNGKVL	83646	12000	315141	7
14751	*Gpi1*	KNLVNKEVM	318694	12000	560091	27
14790	*Grcc10*	SAPENAVRM	3768572	12000	48595185	314
68153	*Gtf2e2*	SGYKFGVL	537587	12000	2451306	45
69237	*Gtpbp4*	QILSDFPKL	395416	12000	225758	33
15016	*H2-Q5*	AMAPRTLLL	5436819	12000	12941105	453
15201	*Hells*	KVLVFSQM	475943	12000	1183394	40
319189	*Hist2h2bb*	SVYVYKVL	871990	12000	3175247	73
382522	*Hist3h2bb*	VNDIFERI	91065	12000	63614	8
15368	*Hmox1*	IQAENAEFM	531189	12000	12000	44
15387	*Hnrpk*	DQIQNAQYL	99408	12000	43071	8
23918	*Impdh2*	GGIQNVGHI	932864	90438	5379787	10
16867	*Lhcgr*	SILLLNAVAF	185982	12000	1613907	15
545270	*LOC545270*	YALYNNWEHM	118954	12000	102363	10
72416	*Lrpprc*	AAIENIEHL	384743	12000	385640	32
99470	*Magi3*	SGLTNRDTL	159958	12000	3688747	13

56692	*Map2k1ip1*	QVVQFNRL	114970	12000	140373	10
26413	*Mapk1*	VGPRYTNL	39166667	149969	45832060	261
212307	*Mapre2*	QLNEQVHSL	493384	12000	1547458	41
17168	*Mare*	SAVKNLQQL	4185921	12000	2766551	349
17169	*Mark3*	TSIAFKNI	57818	12000	289972	5
17427	*Mns1*	KIIEFANI	65858	12000	12000	5
17863	*Myb*	NAIKNHWNSTM	668372	12000	338244	56
67938	*Mylc2b*	SLGKNPTDAYL	503898	12000	4119401	42
74838	*Narg1*	SAVENLNEM	615625	12000	297903	51
17992	*Ndufa4*	VNVDYSKL	14338076	12000	29429606	1195
227197	*Ndufs1*	AKLVNQEVL	503868	12000	908787	42
64652	*Nisch*	TNQDFIQRL	123915	12000	948264	10
68979	*Nol11*	TQLIQTHVL	515769	62667	1499186	8
66394	*Nosip*	TSVRFTQL	3518404	12000	7472080	293
18145	*Npc1*	VAVVNKVDI	243630	12000	356718	20
101706	*Numa1*	VSILNRQVL	365089	12000	799609	30
69482	*Nup35*	AQYGNILKHVM	78184	12000	610648	7
18519	*Pcaf*	VQYKFSHL	7220707	12000	5160299	602
14827	*Pdia3*	FAHTNIESL	7732607	107365	2392144	72
23986	*Peci*	SSYTFPKM	1349403	12000	12000	112
75725	*Phf14*	VNYYFERNM	48515	12000	12000	4
236539	*Phgdh*	TGVVNAQAL	324963	12000	1003626	27
233489	*Picalm*	NGVINAAFM	265420	12000	438191	22
233489	*Picalm*	VAFDFTKV	622109	12000	1026908	52
224938	*Pja2*	SSVQNGIML	186205	33856	525538	5
59047	*Pnkp*	YVHVNRDTL	1160230	12000	178123	97
231329	*Polr2b*	IGPTYYQRL	393502	12000	716661	33
218832	*Polr3a*	VTIQNSELM	206677	12000	12000	17
19015	*Ppard*	ASIVNKDGL	144953	12000	217899	12
170826	*Ppargc1b*	LSVRNGATL	218140	12000	267046	18
54397	*Ppt2*	SSYSFRHL	283225	12000	919892	24
106042	*Prickle1*	LNYKFPGL	376026	12000	124373	31

23992	*Prkra*	ISPENHISL	84113	12000	1206217	7
16913	*Psmb8*	GGVVNMYHM	3322728	12000	185723	277
15170	*Ptpn6*	AQYKFIYV	1303118	12000	1598969	109
78697	*Pus7*	TSIKNQTQL	84865	12000	160379	7
110816	*Pwp2*	FAYRFSNL	218474	12000	126002	18
70314	*Rabep2*	AQVQNSEQL	1384849	46060	843925	30
74734	*Rhoh*	YSVANHNSFL	1605054	82775	2637549	19
68477	*Rmnd5a*	WAVSNREML	79683	12000	116624	7
22644	*Rnf103*	KTIYNVEHL	54697	12000	148600	5
66480	*Rpl15*	STYKFFEV	423190	12000	263479	35
19942	*Rpl27*	KTVVNKDVF	783897	12000	3144986	65
67891	*Rpl4*	VNFVHTNL	2514370	12000	1519392	210
267019	*Rps15a*	VIVRFLTV	514406	12000	12000	43
267019	*Rps15a*	VIVRFLTVM	388796	12000	57863	32
20115	*Rps7*	QQNNVEHKV	1979899	12000	2175109	165
20133	*Rrm1*	FQIVNPHLL	2977570	400671	1830528	7
20133	*Rrm1*	YGIRNSLLI	1677107	174840	647604	10
269254	*Setx*	SQLENKTII	204933	12000	137119	17
81898	*Sf3b1*	KAIVNVIGM	78160	12000	335976	7
170755	*Sgk3*	YSIVNASVL	255048	38491	151335	7
20425	*Shmt1*	RNLDYARL	147317	12000	921297	12
63959	*Slc29a1*	SAIFNNVMTL	61949	12000	295795	5
20527	*Slc2a3*	TGVINAPETI	6643529	12000	5706248	554
20586	*Smarca4*	TQVLNTHYV	276940	12000	121529	23
20740	*Spna2*	KALINADEL	4393266	12000	1485888	366
20815	*Srpk1*	ISGVNGTHI	1109459	12000	1330242	92
67437	*Ssr3*	VAFAYKNV	413665	12000	1416589	34
20448	*St6galnac4*	QVYTFTERM	233788	12000	170957	19
20848	*Stat3*	ATLVFHNL	4023137	12000	6208130	335
58523	*Statip1*	SAPRNFVENF	7865121	248931	13541769	32
16430	*Stt3a*	AVLSFSTRL	42362	12000	41291	4
21687	*Tek*	NDIKFQDV	43651	12000	142042	4

22042	*Tfrc*	KAFTYINL	435591	12000	480537	36
21894	*Tln1*	SVMENSKVLGEAM	522658	12000	352355	44
70549	*Tln2*	SVMENSKVLGESM	52094	12000	41337	4
66074	*Tmem167*	SAIFNFQSL	56289	12000	320506	5
67878	*Tmem33*	QSIAFISRL	394668	12000	166467	33
21916	*Tmod1*	SSIVNKEGL	689429	12000	96959	57
217351	*Tnrc6c*	TSPINPQHM	341556	12000	1238666	28
22142	*Tuba1a*	DIERPTYTNL	656598	157480	1267632	4
22142	*Tuba1a*	ERPTYTNL	816936	86447	698017	9
53382	*Txnl1*	FQNVNSVTL	1199864	12000	1713670	100
67123	*Ubap1*	KSLSFPKL	255902	12000	697677	21
19704	*Upf1*	SGYIYHKL	1545046	12000	391481	129
326622	*Upf2*	SAVIFRTL	244247	12000	1102686	20
216825	*Usp22*	FAVVNHQGTL	711737	12000	12000	59
230895	*Vps13d*	INIHYTQL	188518	12000	90025	16
80743	*Vps16*	VSFTYRYL	794606	57587	689203	14
233405	*Vps33b*	ASLVNADKL	2323671	12000	2376572	194
65114	*Vps35*	FSEENHEPL	345757	12000	562939	29
241627	*Wdr76*	FSPTNPAHL	164815	12000	599612	14
22446	*Xlr3b*	VAAANREVL	1654355	12000	1362836	138
103573	*Xpo1*	TILEFSQNM	561596	12000	2010930	47
68090	*Yif1a*	SGYKYVGM	300842	12000	1171758	25
66980	*Zdhhc6*	SVIKFENL	863986	12000	734576	72
67187	*Zmynd19*	ESYSFEARM	132049	12000	253228	11

Table AII.II: Computed MHC binding affinity of MHC I-associated peptides eluted from EL4 cells.

MHC Ia-associated peptides

GeneID	Gene symbol	Peptide	Restriction size	SYFPEITHI	smm	smm-rank
68523	*1110019N10Rik*	AALENTHLL	Db 9-mer	30	327	1/309
66185	*1110037F02Rik*	RQILNADAM	Db 9-mer	25	13	1/3705
67463	*1200014M14Rik*	GALENAKAEI	Db 10-mer	30	286	2/1099
74174	*1700006H03Rik*	NSPANNIVM	Db 9-mer	25	28	1/317
100910	*2010209O12Rik*	AQIRNLTVL	Db 9-mer	30	20	1/713
76559	*2410024A21Rik*	AGPENSSKI	Db 9-mer	30	2402	110/4133
75425	*2610036D13Rik*	QTVENVEHL	Db 9-mer	27	635	15/2153
67268	*2900073G15Rik*	SMGKNPTDEYL	Db 11-mer	20	344	3/327
338523	*A630082K20Rik*	SSIQNGKYTL	Db 10-mer	29	273	9/1761
18671	*Abcb1a*	STVRNADVI	Db 9-mer	28	10	1/2535
11307	*Abcg1*	VNIEFKDL	Kb 8-mer	23	107	16/1177
104112	*Acly*	SIANFTNV	Kb 8-mer	20	198	33/2167
66713	*Actr2*	YAGSNFPEHI	Db 10-mer	29	386	4/771
67019	*Actr6*	SGYSFTHI	Kb 8-mer	25	39	4/777
226747	*Ahctf1*	TSVENHEFL	Db 9-mer	30	67	1/2235
223774	*Alg12*	AAMLNYTHI	Db 9-mer	28	4	1/753
81702	*Ankrd17*	ASVLNVNHI	Db 9-mer	27	87	5/5189
11765	*Ap1g1*	FALVNGNNI	Db 9-mer	29	23	1/1633
71435	*Arhgap21*	TGVTNRDLI	Db 9-mer	29	392	7/3875
11984	*Atp6v0c*	SAMVFSAM	Kb 8-mer	19	84	3/295
328099	*AU021838*	VSYLFSHV	Kb 8-mer	24	3	1/487
66165	*Bccip*	KAPVNTAEL	Db 9-mer	31	43	1/615
12176	*Bnip3*	LSMRNTSVM	Db 9-mer	24	14	1/357
107976	*Bre*	ATQVYPKL	Kb 8-mer	22	67	7/751
76740	*C920006C10Rik*	SAIINEDNL	Db 9-mer	29	19	1/1621

69719	*Cad*	RQAENGMYI	Db 9-mer	24	732	26/4433
12340	*Capza1*	ISFKFDHL	Kb 8-mer	23	6	1/557
26885	*Casp8ap2*	VTVLNVDHL	Db 9-mer	26	30	1/3907
12449	*Ccnf*	SQAVNKQQI	Db 9-mer	26	588	4/1537
12468	*Cct7*	RAIKNDSVV	Db 9-mer	22	27	1/1071
12576	*Cdkn1b*	FGPVNHEEL	Db 9-mer	33	77	1/377
229841	*Cenpe*	SALQNAESDRL	Db 11-mer	22	76	1/4925
216859	*Centb1*	INQIYEARV	Kb 9-mer	13	268	243/1265
12649	*Chek1*	TGPSNVDKL	Db 9-mer	29	784	7/935
234594	*Cnot1*	IAFIFNNL	Kb 8-mer	23	5	1/4751
12877	*Cpeb1*	SMLQNPLGNVL	Db 11-mer	23	456	10/1105
68975	*Crsp8*	SGLLNQQSL	Db 9-mer	28	118	1/605
71745	*Cul2*	VINSFVHV	Kb 8-mer	17	527	92/1397
215821	*D10Bwg1379e*	AVLTNQETI	Db 9-mer	28	73	4/4323
52635	*D12Ertd551e*	RQLENGTTL	Db 9-mer	28	33	1/1673
218978	*D14Ertd436e*	SQHVNLDQL	Db 9-mer	29	61	1/967
229473	*D930015E06Rik*	ASLVNSPSYL	Db 10-mer	31	730	48/3177
208846	*Daam1*	AAKVNMTEL	Db 9-mer	30	171	4/2137
68087	*Dcakd*	VIQVFQQL	Kb 8-mer	22	33	2/447
67755	*Ddx47*	KTFLFSATM	Kb 9-mer	11	33	14/755
13209	*Ddx6*	INFDFPKL	Kb 8-mer	24	7	1/811
72121	*Dennd2d*	VIYPFMQGL	Kb 9-mer	18	4	4/923
217207	*Dhx8*	QLYTLGAL	Kb 8-mer	16	1953	332/2473
94176	*Dock2*	SMVQNRVFL	Db 9-mer	29	7	2/3639
233115	*Dpy19l3*	SVVAFHNL	Kb 8-mer	22	44	6/1417
76843	*Dtl*	SSPENKNWL	Db 9-mer	29	291	2/1441
13424	*Dync1h1*	ASYEFVQRL	Kb 9-mer	20	18	4/9133
66967	*Edem3*	FMATNPEHL	Db 9-mer	28	32	1/1845
22446	*Xlr3b*	VAAANREVL	Db 9-mer	26	111	1/435
13669	*Eif3s10*	QSIEFSRL	Kb 8-mer	24	28	1/1267
15569	*Elavl2*	GAVTNVKVI	Db 9-mer	26	42	1/703

14791	*Emg1*	VSVEYTEKM	Kb 9-mer	13	58	8/473
13877	*Erh*	YQPYNKDWI	Db 9-mer	26	48	1/191
14050	*Eya3*	TIIIFHSL	Kb 8-mer	22	45	4/1005
212377	*F730047E07Rik*	VGVTYRTL	Kb 8-mer	23	28	2/1911
70611	*Fbxo33*	SSLSNPIANTM	Db 11-mer	17	107	1/575
14137	*Fdft1*	ISLEFRNL	Kb 8-mer	25	3	1/677
57778	*Fmnl1*	LQYEFTHL	Kb 8-mer	29	19	1/2173
22379	*Fmnl3*	LQYEFTKL	Kb 8-mer	30	25	2/2011
239554	*Foxred2*	LGVTNFVHM	Db 9-mer	28	247	3/1313
56784	*Garnl1*	GAIVNGKVL	Db 9-mer	27	27	2/4149
14751	*Gpi1*	KNLVNKEVM	Db 9-mer	27	198	2/1099
14790	*Grcc10*	SAPENAVRM	Db 9-mer	33	165	2/235
68153	*Gtf2e2*	SGYKFGVL	Kb 8-mer	29	18	1/569
69237	*Gtpbp4*	QILSDFPKL	Kb 9-mer	23	231	19/1253
15201	*Hells*	KVLVFSQM	Kb 8-mer	18	254	22/1627
319189	*Hist2h2bb*	SVYVYKVL	Kb 8-mer	27	46	1/237
382522	*Hist3h2bb*	VNDIFERI	Kb 8-mer	18	1748	43/293
15368	*Hmox1*	IQAENAEFM	Db 9-mer	27	308	3/561
15387	*Hnrpk*	DQIQNAQYL	Db 9-mer	28	52	1/909
23918	*Impdh2*	GGIQNVGHI	Db 9-mer	26	291	5/1011
16867	*Lhcgr*	SILLLNAVAF	Db 10-mer	22	2815	564/1383
545270	*LOC545270*	YALYNNWEHM	Db 10-mer	28	45	2/165
72416	*Lrpprc*	AAIENIEHL	Db 9-mer	32	64	4/2595
99470	*Magi3*	SGLTNRDTL	Db 9-mer	31	14	1/2235
56692	*Map2k1ip1*	QVVQFNRL	Kb 8-mer	20	368	14/233
26413	*Mapk1*	VGPRYTNL	Kb 8-mer	25	12	2/701
17168	*Mare*	SAVKNLQQL	Db 9-mer	31	85	1/1121
17169	*Mark3*	TSIAFKNI	Kb 8-mer	19	88	5/1473
17427	*Mns1*	KIIEFANI	Kb 8-mer	20	20	1/967
17863	*Myb*	NAIKNHWNSTM	Db 11-mer	17	309	2/1255
67938	*Mylc2b*	SLGKNPTDAYL	Db 11-mer	16	13242	31/327

74838	*Narg1*	SAVENLNEM	Db 9-mer	33	94	2/1713
17992	*Ndufa4*	VNVDYSKL	Kb 8-mer	24	40	1/149
227197	*Ndufs1*	AKLVNQEVL	Db 9-mer	28	175	1/1437
64652	*Nisch*	TNQDFIQRL	Kb 9-mer	14	358	98/3171
66394	*Nosip*	TSVRFTQL	Kb 8-mer	24	33	1/447
18145	*Npc1*	VAVVNKVDI	Db 9-mer	29	73	3/2537
101706	*Numa1*	VSILNRQVL	Db 9-mer	26	14	1/4171
69482	*Nup35*	AQYGNILKHVM	Db 11-mer	20	1543	42/633
18519	*Pcaf*	VQYKFSHL	Kb 8-mer	28	4	1/1611
14827	*Pdia3*	FAHTNIESL	Db 9-mer	29	19	1/993
23986	*Peci*	SSYTFPKM	Kb 8-mer	25	8	1/561
75725	*Phf14*	VNYYFERNM	Kb 9-mer	18	18	6/1603
236539	*Phgdh*	TGVVNAQAL	Db 9-mer	29	203	2/1049
233489	*Picalm*	NGVINAAFM	Db 9-mer	26	192	4/1303
233489	*Picalm*	VAFDFTKV	Kb 8-mer	20	45	6/1305
224938	*Pja2*	SSVQNGIML	Db 9-mer	26	15	1/1397
59047	*Pnkp*	YVHVNRDTL	Db 9-mer	25	9	1/1027
231329	*Polr2b*	IGPTYYQRL	Kb 9-mer	20	23	15/2319
218832	*Polr3a*	VTIQNSELM	Db 9-mer	27	503	13/2763
19015	*Ppard*	ASIVNKDGL	Db 9-mer	30	117	3/863
170826	*Ppargc1b*	LSVRNGATL	Db 9-mer	27	41	1/2011
54397	*Ppt2*	SSYSFRHL	Kb 8-mer	29	4	1/589
106042	*Prickle1*	LNYKFPGL	Kb 8-mer	29	5	1/1509
23992	*Prkra*	ISPENHISL	Db 9-mer	28	109	2/609
16913	*Psmb8*	GGVVNMYHM	Db 9-mer	27	135	2/535
15170	*Ptpn6*	AQYKFIYV	Kb 8-mer	25	23	1/1039
78697	*Pus7*	TSIKNQTQL	Db 9-mer	28	159	2/1303
110816	*Pwp2*	FAYRFSNL	Kb 8-mer	30	2	1/1823
70314	*Rabep2*	AQVQNSEQL	Db 9-mer	28	382	2/1005
74734	*Rhoh*	YSVANHNSFL	Db 10-mer	26	767	7/365
68477	*Rmnd5a*	WAVSNREML	Db 9-mer	27	158	1/765

22644	*Rnf103*	KTIYNVEHL	Db 9-mer	26	308	4/1349
66480	*Rpl15*	STYKFFEV	Kb 8-mer	24	51	3/393
19942	*Rpl27*	KTVVNKDVF	Db 9-mer	25	54	1/255
67891	*Rpl4*	VNFVHTNL	Kb 8-mer	13	51	1/823
267019	*Rps15a*	VIVRFLTV	Kb 8-mer	19	33	1/245
267019	*Rps15a*	VIVRFLTVM	Kb 9-mer	19	33	9/245
20133	*Rrm1*	FQIVNPHLL	Db 9-mer	28	71	2/1567
20133	*Rrm1*	YGIRNSLLI	Db 9-mer	26	452	7/1567
269254	*Setx*	SQLENKTII	Db 9-mer	28	61	2/5275
81898	*Sf3b1*	KAIVNVIGM	Db 9-mer	28	70	2/2591
170755	*Sgk3*	YSIVNASVL	Db 9-mer	28	6	1/975
20425	*Shmt1*	RNLDYARL	Kb 8-mer	23	22	2/941
63959	*Slc29a1*	SAIFNNVMTL	Db 10-mer	29	886	16/899
20527	*Slc2a3*	TGVINAPETI	Db 10-mer	30	2873	57/969
20586	*Smarca4*	TQVLNTHYV	Db 9-mer	22	94	3/3211
20740	*Spna2*	KALINADEL	Db 9-mer	31	19	1/4817
20815	*Srpk1*	ISGVNGTHI	Db 9-mer	28	55	1/1279
67437	*Ssr3*	VAFAYKNV	Kb 8-mer	18	18	1/355
20448	*St6galnac4*	QVYTFTERM	Kb 9-mer	17	137	13/589
20848	*Stat3*	ATLVFHNL	Kb 8-mer	22	23	1/1429
58523	*Statip1*	SAPRNFVENF	Db 10-mer	30	315	11/1645
16430	*Stt3a*	AVLSFSTRL	Kb 9-mer	13	154	66/1695
21687	*Tek*	NDIKFQDV	Kb 8-mer	18	4583	645/2231
22042	*Tfrc*	KAFTYINL	Kb 8-mer	23	11	2/1371
21894	*Tln1*	SVMENSKVLGEAM	Db 13-mer	23	60	5/5065
70549	*Tln2*	SVMENSKVLGESM	Db 13-mer	23	60	5/5069
66074	*Tmem167*	SAIFNFQSL	Db 9-mer	30	48	1/127
67878	*Tmem33*	QSIAFISRL	Kb 9-mer	13	75	28/479
21916	*Tmod1*	SSIVNKEGL	Db 9-mer	30	228	3/701
217351	*Tnrc6c*	TSPINPQHM	Db 9-mer	28	158	5/3783
22142	*Tuba1a*	ERPTYTNL	Kb 8-mer	23	1759	134/889

22142	*Tuba1a*	DIERPTYTNL	Kb 10-mer	23	849	134/889
53382	*Txnl1*	FQNVNSVTL	Db 9-mer	28	27	1/561
67123	*Ubap1*	KSLSFPKL	Kb 8-mer	22	42	3/867
19704	*Upf1*	SGYIYHKL	Kb 8-mer	29	18	2/2211
326622	*Upf2*	SAVIFRTL	Kb 8-mer	24	57	4/2523
216825	*Usp22*	FAVVNHQGTL	Db 10-mer	31	985	7/1033
230895	*Vps13d*	INIHYTQL	Kb 8-mer	25	39	10/8753
80743	*Vps16*	VSFTYRYL	Kb 8-mer	22	3	1/1663
233405	*Vps33b*	ASLVNADKL	Db 9-mer	31	25	1/1217
65114	*Vps35*	FSEENHEPL	Db 9-mer	27	33	3/1575
241627	*Wdr76*	FSPTNPAHL	Db 9-mer	29	18	1/1031
103573	*Xpo1*	TILEFSQNM	Kb 9-mer	14	88	81/2127
68090	*Yif1a*	SGYKYVGM	Kb 8-mer	25	17	1/571
66980	*Zdhhc6*	SVIKFENL	Kb 8-mer	25	44	3/811
67187	*Zmynd19*	ESYSFEARM	Kb 9-mer	18	386	2/439

MHC Ib-associated peptides

GeneID	Gene symbol	Peptide	Restriction size	Rankpep
77877	*6030458C11Rik*	KVLDVLHSL	Qa2 9-mer	228
213895	*Bms1l*	YQNQEIHNL	Qa2 9-mer	131
12181	*Bop1*	EQVALVHRL	Qa2 9-mer	199
224171	*C330027C09Rik*	AQNDIEHLF	Qa2 9-mer	155
12465	*Cct5*	SMMDVDHQI	Qa2 9-mer	223
214901	*Chtf18*	QLSDTLHSL	Qa2 9-mer	184
80986	*Ckap2*	TLNDLIHNI	Qa2 9-mer	192
12847	*Copa*	SQLPVDHIL	Qa2 9-mer	196
66366	*Ergic3*	QQLDVEHNL	Qa2 9-mer	212
108655	*Foxp1*	QQLQQQHLL	Qa2 9-mer	148
15016	*H2-Q5*	AMAPRTLLL	Qa1 9-mer	
212307	*Mapre2*	QLNEQVHSL	Qa2 9-mer	176

68979	*Nol11*	TQLIQTHVL	Qa2 9-mer	150
20115	*Rps7*	QQNNVEHKV	Qa2 9-mer	180

Table AII.III: Computed MHC binding affinity of MHC I-associated peptides from primary mouse thymocytes.

MHC Ia-associated peptides

GeneID	Gene symbol	Peptides	Restriction size	smm	SYFPEITHI
68295	*0610011L14Rik*	ISILYHQL	Kb 8-mer	20	24
68523	*1110019N10Rik*	AALENTHLL	Db 9-mer	327	30
72368	*2310045N01Rik*	SSVYFRSV	Kb 8-mer	51	19
75425	*2610036D13Rik*	QTVENVEHL	Db 9-mer	635	27
72722	*2810405J04Rik*	ASPEFTKL	Kb 8-mer	60	26
72747	*2810439F02Rik*	SGYDFENRL	Kb 9-mer	19	21
67268	*2900073G15Rik*	SMGKNPTDEYL	Db 11-mer	344	20
67392	*4833420G17Rik*	FVISNYREQL	Db 11-mer	1238	26
218989	*6720456H20Rik*	TGIRNLEWL	Db 9-mer	574	30
11307	*Abcg1*	VNIEFKDL	Kb 8-mer	107	23
66713	*Actr2*	YAGSNFPEHI	Db 10-mer	386	29
67019	*Actr6*	SGYSFTHI	Kb 8-mer	39	25
57869	*Adck2*	SGPTYIKL	Kb 8-mer	78	25
81702	*Ankrd17*	ASVLNVNHI	Db 9-mer	87	27
329154	*Ankrd44*	ASVINGHTL	Db 9-mer	60	28
73341	*Arhgef6*	AIIAFKTL	Kb 8-mer	45	24
54208	*Arl6ip1*	IILKYIGM	Kb 8-mer	11	20
192195	*Ash1l*	VGLINKDSV	Db 9-mer	412	24
11906	*Atbf1*	VAIGNPVHL	Db 9-mer	292	28
Q9JKK8	*Atr*	FQALNAEKL	Db 9-mer	208	26
54138	*Atxn10*	NSIRNLDTI	Db 9-mer	13	31
328099	*AU021838*	VSYLFSHV	Kb 8-mer	3	24
213550	*AV340375*	FSNKNLEEL	Db 9-mer	53	31
235461	*B230380D07Rik*	VNVRFTGV	Kb 8-mer	36	19
224727	*Bat3*	SIAAFIQRL	Kb 9-mer	199	14

66165	*Bccip*	KAPVNTAEL	Db 9-mer	43	31
107976	*Bre*	ATQVYPKL	Kb 8-mer	67	22
26885	*Casp8ap2*	VTVLNVDHL	Db 9-mer	30	26
76041	*Ccdc125*	NAVLNQRYL	Db 9-mer	32	27
77048	*Ccdc41*	AQVENVQRI	Db 9-mer	498	29
12445	*Ccnd3*	AAVIAHDFL	Db 9-mer	46	20
12449	*Ccnf*	SQAVNKQQI	Db 9-mer	588	26
12450	*Ccng1*	TAFQFLQL	Kb 8-mer	80	22
12501	*Cd3e*	YSGLNQRAV	Db 9-mer	406	19
26364	*Cd97*	KLLSNINSVF	Db 10-mer	9292	23
12576	*Cdkn1b*	FGPVNHEEL	Db 9-mer	77	33
229841	*Cenpe*	SALQNAESDRL	Db 12-mer	76	22
216859	*Centb1*	INQIYEARV	Kb 9-mer	268	13
12649	*Chek1*	TGPSNVDKL	Db 9-mer	784	29
243764	*Chrm2*	TVIGYWPL	Kb 8-mer	196	21
70349	*Copb1*	IALRYVAL	Kb 8-mer	22	23
67876	*Coq10b*	ISFEFRSL	Kb 8-mer	2	24
68975	*Crsp8*	SGLLNQQSL	Db 9-mer	118	28
71745	*Cul2*	VINSFVHV	Kb 8-mer	527	17
12767	*Cxcr4*	VVFQFQHI	Kb 8-mer	56	16
52635	*D12Ertd551e*	RQLENGTTL	Db 9-mer	33	28
218978	*D14Ertd436e*	SQHVNLDQL	Db 9-mer	61	29
71919	*D15Ertd682e*	TNVLFNHL	Kb 8-mer	169	21
69601	*Dab2ip*	VSPTNPTKL	Db 9-mer	98	29
67755	*Ddx47*	KTFLFSATM	Kb 9-mer	33	11
13207	*Ddx5*	NQAINPKLLQL	Db 11-mer	4239	22
13209	*Ddx6*	INFDFPKL	Kb 8-mer	7	24
76131	*Depdc1a*	FSQENTEKI	Db 9-mer	231	28
13433	*Dnmt1*	LSLENGTHTL	Db 10-mer	758	28
94176	*Dock2*	SMVQNRVFL	Db 9-mer	7	29
233115	*Dpy19l3*	SVVAFHNL	Kb 8-mer	44	22
13424	*Dync1h1*	ASYEFVQRL	Kb 9-mer	18	20

13669	*Eif3s10*	QSIEFSRL	Kb 8-mer	28	24
73830	*Eif3s12*	VGITYQHI	Kb 8-mer	100	18
170439	*Elovl6*	KSFLFSAL	Kb 8-mer	10	22
10436	*EMG1*	VSVEYTEKM	Kb 9-mer	58	13
13819	*Epas1*	SNYLFTKL	Kb 8-mer	10	30
212377	*F730047E07Rik*	VGVTYRTL	Kb 8-mer	28	23
107035	*Fbxo38*	SAFSFRTL	Kb 8-mer	22	24
14137	*Fdft1*	ISLEFRNL	Kb 8-mer	3	25
57778	*Fmnl1*	LQYEFTHL	Kb 8-mer	19	29
22379	*Fmnl3*	LQYEFTKL	Kb 8-mer	25	30
14359	*Fxr1h*	EIVTFERL	Kb 8-mer	464	22
14790	*Grcc10*	SAPENAVRM	Db 9-mer	165	33
68153	*Gtf2e2*	SGYKFGVL	Kb 8-mer	18	29
69237	*Gtpbp4*	QILSDFPKL	Kb 9-mer	231	23
14976	*H2-Ke2*	SNVVFKLL	Kb 9-mer	145	22
15019	*H2-Q8*	FAYEGRDYI	Db 9-mer	12	21
15201	*Hells*	KVLVFSQM	Kb 8-mer	254	18
15257	*Hipk1*	SLLTNHVTL	Db 9-mer	167	28
319189	*Hist2h2bb*	SVYVYKVL	Kb 8-mer	46	27
98758	*Hnrpf*	FSPLNPVRV	Db 9-mer	127	25
15441	*Hp1bp3*	VSQYYPKL	Kb 8-mer	29	22
16145	*Igtp*	IVAENTKTSL	Db 10-mer	3059	24
23918	*Impdh2*	GGIQNVGHI	Db 9-mer	291	26
16329	*Inpp1*	YMGTNIHSL	Db 9-mer	84	25
76582	*Ipo11*	TIILFTKV	Kb 8-mer	95	20
338523	*Jhdm1d*	SSIQNGKYTL	Db 10-mer	273	29
16211	*Kpnb1*	AALQNLVKI	Db 9-mer	77	31
110308	*Krt5*	AAYMNKVEL	Db 9-mer	63	28
110310	*Krt7*	AAYTNKVEL	Db 9-mer	66	28
72416	*Lrpprc*	AAIENIEHL	Db 9-mer	64	32
26413	*Mapk1*	VGPRYTNL	Kb 8-mer	12	25
17168	*Mare*	SAVKNLQQL	Db 9-mer	85	31

13728	*Mark2*	ASIQNGKDSL	Db 10-mer	1344	30
217615	*Mgea6*	RNQVYTQL	Kb 8-mer	224	24
67014	*Mina*	VVYIYHSL	Kb 8-mer	9	26
17427	*Mns1*	KIIEFANI	Kb 8-mer	20	20
121022	*Mrps6*	SAVENILEHL	Db 10-mer	178	32
219135	*Mtmr6*	VGIENIHVM	Db 9-mer	150	27
17886	*Myh9*	RVAEFTTNL	Kb 9-mer	228	14
67938	*Mylc2b*	SLGKNPTDAYL	Db 11-mer	344	16
67608	*Narf*	AAYGFRNI	Kb 8-mer	32	26
17992	*Ndufa4*	VNVDYSKL	Kb 8-mer	40	24
227197	*Ndufs1*	AKLVNQEVL	Db 9-mer	175	28
18034	*Nfkb2*	AALQNLEQL	Db 9-mer	98	32
64652	*Nisch*	TNQDFIQRL	Kb 9-mer	358	14
97895	*Nlrp4f*	TPVKNIDTV	Db 9-mer	509	24
68979	*Nol11*	IAVSFREL	Kb 8-mer	48	23
211548	*Nomo1*	SSLVNKEDV	Db 9-mer	99	26
66394	*Nosip*	TSVRFTQL	Kb 8-mer	33	24
18145	*Npc1*	VAVVNKVDI	Db 9-mer	73	29
101706	*Numa1*	VSILNRQVL	Db 9-mer	14	26
70699	*Nup205*	VNNEFEKL	Kb 8-mer	1138	24
66923	*Pbrm1*	SQVYNDAHI	Db 9-mer	111	26
18538	*Pcna*	RAEDNADTLAL	Db 11-mer	59	27
56426	*Pdcd10*	ILQTFKTVA	Kb 9-mer	5	18
207728	*Pde2a*	IKNENQEVI	Db 9-mer	1260	24
23986	*Peci*	SSYTFPKM	Kb 8-mer	8	25
73739	*Pgea1*	ASLSNLHSL	Db 9-mer	218	30
233489	*Picalm*	VAFDFTKV	Kb 8-mer	45	20
83490	*Pik3ap1*	YGLKNLTAL	Db 9-mer	40	29
18950	*Pnp*	SLITNKVVM	Db 9-mer	104	26
231329	*Polr2b*	IGPTYYQRL	Kb 9-mer	23	20
66385	*Ppp1r7*	RAIENIDTL	Db 9-mer	7	30
51792	*Ppp2r1a*	IIPMFSNL	Kb 8-mer	4	25

73699	*Ppp2r1b*	VADKFSEL	Kb 8-mer	375	23
214572	*Prmt7*	VIVEFRDL	Kb 8-mer	28	23
19167	*Psma3*	KAVENSSTAI	Db 10-mer	517	30
16913	*Psmb8*	GGVVNMYHM	Db 9-mer	135	27
15170	*Ptpn6*	AQYKFIYV	Kb 8-mer	23	25
19696	*Rel*	TTLIFQKL	Kb 8-mer	58	21
26374	*Rfwd2*	YVVDNIDHL	Db 9-mer	156	25
74734	*Rhoh*	YSVANHNSFL	Db 10-mer	767	26
68477	*Rmnd5a*	WAVSNREML	Db 9-mer	158	27
20054	*Rps15*	VGVYNGKTF	Db 9-mer	3132	25
267019	*Rps15a*	ISPRFDVQL	Kb 9-mer	172	14
267019	*Rps15a*	VIVRFLTV	Kb 8-mer	33	19
20115	*Rps7*	VNFEFPEF	Kb 8-mer	24	12
20133	*Rrm1*	FQIVNPHLL	Db 9-mer	71	28
20133	*Rrm1*	YGIRNSLLI	Db 9-mer	452	26
235623	*Scap*	SGYDFSRL	Kb 8-mer	5	30
13722	*Scye1*	SGLVNHVPL	Db 9-mer	59	29
20338	*Sell1*	QALKYFNL	Kb 8-mer	171	23
20393	*Sgk*	STLTYSRM	Kb 8-mer	38	18
20740	*Spna2*	KALINADEL	Db 9-mer	19	31
20815	*Srpk1*	ISGVNGTHI	Db 9-mer	55	28
20443	*St3gal4*	GAVKNLTYF	Db 9-mer	81	30
20848	*Stat3*	ATLVFHNL	Kb 8-mer	23	22
71728	*Stk11ip*	SALRFLNL	Kb 8-mer	23	24
16430	*Stt3a*	AVLSFSTRL	Kb 9-mer	154	13
108143	*Taf9*	SGLKYVNV	Kb 8-mer	45	20
216198	*Tcp11l2*	KATTNIVEM	Db 9-mer	225	27
217353	*Tmc6*	ASYLFRGL	Kb 8-mer	3	29
67511	*Tmed9*	VIGNYRTQL	Kb 9-mer	271	14
72759	*Tmem135*	KGFTFSAL	Kb 8-mer	23	22
66074	*Tmem167*	AIFNFQSL	Kb 8-mer	8	22
67878	*Tmem33*	QSIAFISRL	Kb 9-mer	75	13

21916	*Tmod1*	SSIVNKEGL	Db 9-mer	228	30
22057	*Tob1*	ISYLYNKL	Kb 8-mer	4	29
21990	*Tph1*	FSLENEVGGL	Db 10-mer	728	28
21990	*Tph1*	WGTIFREL	Kb 8-mer	309	22
12751	*Tpp1*	IQRVNTEFM	Db 9-mer	100	25
100683	*Trrap*	QDPVFQKL	Kb 8-mer	885	22
73710	*Tubb2b*	VNMVPFPRL	Kb 9-mer	252	20
56085	*Ubqln1*	RALSNLESI	Db 9-mer	87	30
224826	*Ubr2*	VAYKFPELL	Kb 9-mer	6	28
59025	*Usp14*	VKYLFTGL	Kb 8-mer	20	27
216825	*Usp22*	FAVVNHQGTL	Db 10-mer	985	31
271564	*Vps13a*	VSIQFYHL	Kb 8-mer	3	22
230895	*Vps13d*	INIHYTQL	Kb 8-mer	39	25
80743	*Vps16*	VSFTYRYL	Kb 8-mer	3	22
77573	*Vps33a*	GSLANHTSI	Db 9-mer	161	27
233405	*Vps33b*	ASLVNADKL	Db 9-mer	25	31
65114	*Vps35*	FSEENHEPL	Db 9-mer	33	27
22375	*Wars*	YTVENAKDII	Db 10-mer	492	26
69544	*Wdr5b*	AALENDKTI	Db 9-mer	56	31
74781	*Wipi2*	SGYKFFSL	Kb 8-mer	14	29
68090	*Yif1a*	SGYKYVGM	Kb 8-mer	17	25
66980	*Zdhhc6*	SVIKFENL	Kb 8-mer	44	25

MHC Ib-associated peptides

GeneID	**Gene symbol**	**Peptides**	**Restriction size**	**Rankpep**
319236	*9230105E10Rik*	FISDVEHQL	Qa2 9-mer	181
12014	*Bach2*	EQLEFIHDI	Qa2 9-mer	183
224171	*C330027C09Rik*	AQNDIEHLF	Qa2 9-mer	155
75565	*Ccdc101*	ELLTELHQL	Qa2 9-mer	198
13204	*Dhx15*	TLLNVYHAF	Qa2 9-mer	215
66366	*Ergic3*	QQLDVEHNL	Qa2 9-mer	212

108655	*Foxp1*	QQLQQQHLL	Qa2 9-mer	148
15016	*H2-Q5*	AMAPRTLLL	Qa1 9-mer	
212307	*Mapre2*	QLNEQVHSL	Qa2 9-mer	176
68979	*Nol11*	TQLIQTHVL	Qa2 9-mer	150
69608	*Sec24d*	ESQSVIHNL	Qa2 9-mer	185
319565	*Syne2*	DLIQTIHEL	Qa2 9-mer	213
50875	*Tmod3*	AAILGMHNL	Qa2 9-mer	134
320938	*Tnpo3*	GLLEIAHSL	Qa2 9-mer	237

Table AII.IV: Computed MHC binding affinity of MHC I-associated peptides in vivo grown EL4 cells.

MHC Ia-associated peptides

GeneID	Gene symbol	Peptides	Restriction size	smm	SYFPEITHI
68295	*0610011L14Rik*	ISILYHQL	Kb 8-mer	20	24
68523	*1110019N10Rik*	AALENTHLL	Db 9-mer	327	30
72368	*2310045N01Rik*	SSVYFRSV	Kb 8-mer	51	19
75425	*2610036D13Rik*	QTVENVEHL	Db 9-mer	635	27
72722	*2810405J04Rik*	ASPEFTKL	Kb 8-mer	60	26
72747	*2810439F02Rik*	SGYDFENRL	Kb 9-mer	19	21
67268	*2900073G15Rik*	SMGKNPTDEYL	Db 11-mer	344	20
67392	*4833420G17Rik*	FVISNYREQL	Db 11-mer	1238	26
218989	*6720456H20Rik*	TGIRNLEWL	Db 9-mer	574	30
11307	*Abcg1*	VNIEFKDL	Kb 8-mer	107	23
66713	*Actr2*	YAGSNFPEHI	Db 10-mer	386	29
67019	*Actr6*	SGYSFTHI	Kb 8-mer	39	25
57869	*Adck2*	SGPTYIKL	Kb 8-mer	78	25
81702	*Ankrd17*	ASVLNVNHI	Db 9-mer	87	27
329154	*Ankrd44*	ASVINGHTL	Db 9-mer	60	28
75415	*Arhgap12*	YGLLNVTKI	Db 9-mer	60	26
73341	*Arhgef6*	AIIAFKTL	Kb 8-mer	45	24
54208	*Arl6ip1*	IILKYIGM	Kb 8-mer	11	20
192195	*Ash1l*	VGLINKDSV	Db 9-mer	412	24
11906	*Atbf1*	VAIGNPVHL	Db 9-mer	292	28
Q9JKK8	*Atr*	FQALNAEKL	Db 9-mer	208	26
54138	*Atxn10*	NSIRNLDTI	Db 9-mer	13	31
328099	*AU021838*	VSYLFSHV	Kb 8-mer	3	24
213550	*AV340375*	FSNKNLEEL	Db 9-mer	53	31
235461	*B230380D07Rik*	VNVRFTGV	Kb 8-mer	36	19
224727	*Bat3*	SIAAFIQRL	Kb 9-mer	199	14

66165	*Bccip*	KAPVNTAEL	Db 9-mer	43	31
107976	*Bre*	ATQVYPKL	Kb 8-mer	67	22
26885	*Casp8ap2*	VTVLNVDHL	Db 9-mer	30	26
76041	*Ccdc125*	NAVLNQRYL	Db 9-mer	32	27
77048	*Ccdc41*	AQVENVQRI	Db 9-mer	498	29
12445	*Ccnd3*	AAVIAHDFL	Db 9-mer	46	20
12449	*Ccnf*	SQAVNKQQI	Db 9-mer	588	26
12450	*Ccng1*	TAFQFLQL	Kb 8-mer	80	22
12501	*Cd3e*	YSGLNQRAV	Db 9-mer	406	19
26364	*Cd97*	KLLSNINSVF	Db 10-mer	9292	23
12576	*Cdkn1b*	FGPVNHEEL	Db 9-mer	77	33
229841	*Cenpe*	SALQNAESDRL	Db 12-mer	76	22
216859	*Centb1*	INQIYEARV	Kb 9-mer	268	13
12649	*Chek1*	TGPSNVDKL	Db 9-mer	784	29
243764	*Chrm2*	TVIGYWPL	Kb 8-mer	196	21
70349	*Copb1*	IALRYVAL	Kb 8-mer	22	23
67876	*Coq10b*	ISFEFRSL	Kb 8-mer	2	24
68975	*Crsp8*	SGLLNQQSL	Db 9-mer	118	28
71745	*Cul2*	VINSFVHV	Kb 8-mer	527	17
12767	*Cxcr4*	VVFQFQHI	Kb 8-mer	56	16
52635	*D12Ertd551e*	RQLENGTTL	Db 9-mer	33	28
218978	*D14Ertd436e*	SQHVNLDQL	Db 9-mer	61	29
71919	*D15Ertd682e*	TNVLFNHL	Kb 8-mer	169	21
69601	*Dab2ip*	VSPTNPTKL	Db 9-mer	98	29
67755	*Ddx47*	KTFLFSATM	Kb 9-mer	33	11
13209	*Ddx6*	INFDFPKL	Kb 8-mer	7	24
76131	*Depdc1a*	FSQENTEKI	Db 9-mer	231	28
13433	*Dnmt1*	LSLENGTHTL	Db 10-mer	758	28
94176	*Dock2*	SMVQNRVFL	Db 9-mer	7	29
233115	*Dpy19l3*	SVVAFHNL	Kb 8-mer	44	22
13424	*Dync1h1*	ASYEFVQRL	Kb 9-mer	18	20
13669	*Eif3s10*	QSIEFSRL	Kb 8-mer	28	24

73830	*Eif3s12*	VGITYQHI	Kb 8-mer	100	18
54709	*Eif3s2*	FGPINSVAF	Db 9-mer	1976	26
170439	*Elovl6*	KSFLFSAL	Kb 8-mer	10	22
10436	*EMG1*	VSVEYTEKM	Kb 9-mer	58	13
13819	*Epas1*	SNYLFTKL	Kb 8-mer	10	30
212377	*F730047E07Rik*	VGVTYRTL	Kb 8-mer	28	23
14104	*Fasn*	TALENLSTL	Db 9-mer	48	33
107035	*Fbxo38*	SAFSFRTL	Kb 8-mer	22	24
14137	*Fdft1*	ISLEFRNL	Kb 8-mer	3	25
57778	*Fmnl1*	LQYEFTHL	Kb 8-mer	19	29
22379	*Fmnl3*	LQYEFTKL	Kb 8-mer	25	30
14359	*Fxr1h*	EIVTFERL	Kb 8-mer	464	22
14790	*Grcc10*	SAPENAVRM	Db 9-mer	165	33
68153	*Gtf2e2*	SGYKFGVL	Kb 8-mer	18	29
69237	*Gtpbp4*	QILSDFPKL	Kb 9-mer	231	23
14976	*H2-Ke2*	SNVVFKLL	Kb 9-mer	145	22
15019	*H2-Q8*	FAYEGRDYI	Db 9-mer	12	21
15201	*Hells*	KVLVFSQM	Kb 8-mer	254	18
15257	*Hipk1*	SLLTNHVTL	Db 9-mer	167	28
319189	*Hist2h2bb*	SVYVYKVL	Kb 8-mer	46	27
20585	*Hltf*	SGFVFTRL	Kb 8-mer	7	23
98758	*Hnrpf*	FSPLNPVRV	Db 9-mer	127	25
15441	*Hp1bp3*	VSQYYPKL	Kb 8-mer	29	22
23918	*Impdh2*	GGIQNVGHI	Db 9-mer	291	26
16329	*Inpp1*	YMGTNIHSL	Db 9-mer	84	25
76582	*Ipo11*	TIILFTKV	Kb 8-mer	95	20
338523	*Jhdm1d*	SSIQNGKYTL	Db 10-mer	273	29
16211	*Kpnb1*	AALQNLVKI	Db 9-mer	77	31
76130	*Las1l*	ALVRFVNL	Kb 8-mer	74	23
72416	*Lrpprc*	AAIENIEHL	Db 9-mer	64	32
26413	*Mapk1*	VGPRYTNL	Kb 8-mer	12	25
17168	*Mare*	SAVKNLQQL	Db 9-mer	85	31

13728	*Mark2*	ASIQNGKDSL	Db 10-mer	1344	30
217615	*Mgea6*	RNQVYTQL	Kb 8-mer	224	24
67014	*Mina*	VVYIYHSL	Kb 8-mer	9	26
121022	*Mrps6*	SAVENILEHL	Db 10-mer	178	32
219135	*Mtmr6*	VGIENIHVM	Db 9-mer	150	27
17886	*Myh9*	RVAEFTTNL	Kb 9-mer	228	14
67938	*Mylc2b*	SLGKNPTDAYL	Db 11-mer	344	16
67608	*Narf*	AAYGFRNI	Kb 8-mer	32	26
67563	*Narfl*	VAYGFRNI	Kb 8-mer	15	25
17992	*Ndufa4*	VNVDYSKL	Kb 8-mer	40	24
227197	*Ndufs1*	AKLVNQEVL	Db 9-mer	175	28
18034	*Nfkb2*	AALQNLEQL	Db 9-mer	98	32
64652	*Nisch*	TNQDFIQRL	Kb 9-mer	358	14
97895	*Nlrp4f*	TPVKNIDTV	Db 9-mer	509	24
68979	*Nol11*	IAVSFREL	Kb 8-mer	48	23
211548	*Nomo1*	SSLVNKEDV	Db 9-mer	99	26
66394	*Nosip*	TSVRFTQL	Kb 8-mer	33	24
18145	*Npc1*	VAVVNKVDI	Db 9-mer	73	29
101706	*Numa1*	VSILNRQVL	Db 9-mer	14	26
70699	*Nup205*	VNNEFEKL	Kb 8-mer	1138	24
18813	*Pa2g4*	AQFKFTVL	Kb 8-mer	59	23
66923	*Pbrm1*	SQVYNDAHI	Db 9-mer	111	26
18538	*Pcna*	RAEDNADTLAL	Db 11-mer	59	27
56426	*Pdcd10*	ILQTFKTVA	Kb 9-mer	5	18
207728	*Pde2a*	IKNENQEVI	Db 9-mer	1260	24
23986	*Peci*	SSYTFPKM	Kb 8-mer	8	25
18626	*Per1*	YTLRNQDTF	Db 9-mer	55	25
56612	*Pfdn5*	SMYVPGKL	Kb 8-mer	293	18
73739	*Pgea1*	ASLSNLHSL	Db 9-mer	218	30
233489	*Picalm*	VAFDFTKV	Kb 8-mer	45	20
83490	*Pik3ap1*	YGLKNLTAL	Db 9-mer	40	29
18950	*Pnp*	SLITNKVVM	Db 9-mer	104	26

231329	*Polr2b*	IGPTYYQRL	Kb 9-mer	23	20
18985	*Pou2af1*	SGPQFVQL	Kb 8-mer	172	24
66385	*Ppp1r7*	RAIENIDTL	Db 9-mer	7	30
51792	*Ppp2r1a*	IIPMFSNL	Kb 8-mer	4	25
73699	*Ppp2r1b*	VADKFSEL	Kb 8-mer	375	23
106042	*Prickle1*	LNYKFPGL	Kb 8-mer	5	29
214572	*Prmt7*	VIVEFRDL	Kb 8-mer	28	23
19167	*Psma3*	KAVENSSTAI	Db 10-mer	517	30
16913	*Psmb8*	GGVVNMYHM	Db 9-mer	135	27
15170	*Ptpn6*	AQYKFIYV	Kb 8-mer	23	25
108911	*Rcc2*	AAYRNLGQNL	Db 10-mer	924	29
19696	*Rel*	TTLIFQKL	Kb 8-mer	58	21
26374	*Rfwd2*	YVVDNIDHL	Db 9-mer	156	25
74734	*Rhoh*	YSVANHNSFL	Db 10-mer	767	26
20054	*Rps15*	VGVYNGKTF	Db 9-mer	3132	25
267019	*Rps15a*	VIVRFLTV	Kb 8-mer	33	19
267019	*Rps15a*	ISPRFDVQL	Kb 9-mer	172	14
20115	*Rps7*	VNFEFPEF	Kb 8-mer	24	12
20133	*Rrm1*	FQIVNPHLL	Db 9-mer	71	28
20133	*Rrm1*	YGIRNSLLI	Db 9-mer	452	26
235623	*Scap*	SGYDFSRL	Kb 8-mer	5	30
13722	*Scye1*	SGLVNHVPL	Db 9-mer	59	29
20338	*Sell1*	QALKYFNL	Kb 8-mer	171	23
20393	*Sgk*	STLTYSRM	Kb 8-mer	38	18
432572	*Specc1*	TSLAFESRL	Kb 9-mer	97	12
20740	*Spna2*	KALINADEL	Db 9-mer	19	31
20815	*Srpk1*	ISGVNGTHI	Db 9-mer	55	28
20443	*St3gal4*	GAVKNLTYF	Db 9-mer	81	30
20848	*Stat3*	ATLVFHNL	Kb 8-mer	23	22
71728	*Stk11ip*	SALRFLNL	Kb 8-mer	23	24
16430	*Stt3a*	AVLSFSTRL	Kb 9-mer	154	13
108143	*Taf9*	SGLKYVNV	Kb 8-mer	45	20

245638	*Tbc1d8b*	TAFRFSEL	Kb 8-mer	23	23
216198	*Tcp11l2*	KATTNIVEM	Db 9-mer	225	27
217353	*Tmc6*	ASYLFRGL	Kb 8-mer	3	29
72759	*Tmem135*	KGFTFSAL	Kb 8-mer	23	22
66074	*Tmem167*	AIFNFQSL	Kb 8-mer	8	22
67878	*Tmem33*	QSIAFISRL	Kb 9-mer	75	13
21916	*Tmod1*	SSIVNKEGL	Db 9-mer	228	30
22057	*Tob1*	ISYLYNKL	Kb 8-mer	4	29
21973	*Top2a*	NSMVLFDHV	Db 9-mer	228	14
21990	*Tph1*	FSLENEVGGL	Db 10-mer	728	28
21990	*Tph1*	WGTIFREL	Kb 8-mer	309	22
12751	*Tpp1*	IQRVNTEFM	Db 9-mer	100	25
100683	*Trrap*	QDPVFQKL	Kb 8-mer	885	22
73710	*Tubb2b*	VNMVPFPRL	Kb 9-mer	252	20
56085	*Ubqln1*	RALSNLESI	Db 9-mer	87	30
224826	*Ubr2*	VAYKFPELL	Kb 9-mer	6	28
59025	*Usp14*	VKYLFTGL	Kb 8-mer	20	27
216825	*Usp22*	FAVVNHQGTL	Db 10-mer	985	31
271564	*Vps13a*	VSIQFYHL	Kb 8-mer	3	22
230895	*Vps13d*	INIHYTQL	Kb 8-mer	39	25
80743	*Vps16*	VSFTYRYL	Kb 8-mer	3	22
77573	*Vps33a*	GSLANHTSI	Db 9-mer	161	27
233405	*Vps33b*	ASLVNADKL	Db 9-mer	25	31
65114	*Vps35*	FSEENHEPL	Db 9-mer	33	27
218035	*Vps41*	IGLAYVNHL	Kb 9-mer	17	13
22375	*Wars*	YTVENAKDII	Db 10-mer	492	26
69544	*Wdr5b*	AALENDKTI	Db 9-mer	56	31
74781	*Wipi2*	SGYKFFSL	Kb 8-mer	14	29
22446	*Xlr3b*	VAAANREVL	Db 9-mer	111	26
68090	*Yif1a*	SGYKYVGM	Kb 8-mer	17	25
66980	*Zdhhc6*	SVIKFENL	Kb 8-mer	44	25

MHC Ib-associated peptides

GeneID	Gene symbol	Peptides	Restriction size	Rankpep
224171	*C330027C09Rik*	AQNDIEHLF	Qa2 9-mer	155
75565	*Ccdc101*	ELLTELHQL	Qa2 9-mer	198
13204	*Dhx15*	TLLNVYHAF	Qa2 9-mer	215
66366	*Ergic3*	QQLDVEHNL	Qa2 9-mer	212
108655	*Foxp1*	QQLQQQHLL	Qa2 9-mer	148
15016	*H2-Q5*	AMAPRTLLL	Qa1 9-mer	
212307	*Mapre2*	QLNEQVHSL	Qa2 9-mer	176
68979	*Nol11*	TQLIQTHVL	Qa2 9-mer	150
69608	*Sec24d*	ESQSVIHNL	Qa2 9-mer	185
319565	*Syne2*	DLIQTIHEL	Qa2 9-mer	213
50875	*Tmod3*	AAILGMHNL	Qa2 9-mer	134

Table AII.V: Relative abundance of MHC I peptides eluted from primary thymocytes versus in vivo grown EL4.

GeneID	Gene symbol	Peptide	Average intensities from EL4	Thymocytes	p-values	EL4/Thy
68295	*0610011L14Rik*	ISILYHQL	810647	1248361	0.10	-1.5
68523	*1110019N10Rik*	AALENTHLL	116372	145359	0.63	-1.2
72368	*2310045N01Rik*	SSVYFRSV	94609	42165	0.09	2.2
75425	*2610036D13Rik*	QTVENVEHL	268223	422026	0.20	-1.6
72722	*2810405J04Rik*	ASPEFTKL	1928424	879573	0.07	2.2
72747	*2810439F02Rik*	SGYDFENRL	55031	123324	0.13	-2.2
67268	*2900073G15Rik*	SMGKNPTDEYL	2911106	762807	0.05	3.8
218989	*6720456H20Rik*	TGIRNLEWL	197157	149305	0.56	1.3
319236	*9230105E10Rik*	FISDVEHQL	12000	278323	0.01	-23.2
11307	*Abcg1*	VNIEFKDL	2074080	1805346	0.54	1.1
66713	*Actr2*	YAGSNFPEHI	2319817	1277449	0.04	1.8
67019	*Actr6*	SGYSFTHI	436804	141915	0.06	3.1
57869	*Adck2*	SGPTYIKL	86469	161672	0.26	-1.9
81702	*Ankrd17*	ASVLNVNHI	5882733	5485733	0.65	1.1
329154	*Ankrd44*	ASVINGHTL	385034	353775	0.79	1.1
54208	*Arl6ip1*	IILKYIGM	246911	127103	0.16	1.9
192195	*Ash1l*	VGLINKDSV	241257	132489	0.28	1.8
Q9JKK8	*Atr*	FQALNAEKL	60707	89418	0.20	-1.5
54138	*Atxn10*	NSIRNLDTI	11787170	22819844	0.03	-1.9
328099	*AU021838*	VSYLFSHV	432991	356073	0.63	1.2
235461	*B230380D07Rik*	VNVRFTGV	54488	74969	0.31	-1.4
12014	*Bach2*	EQLEFIHDI	12000	242371	0.01	-20.2
224727	*Bat3*	SIAAFIQRL	133395	93668	0.45	1.4
66165	*Bccip*	KAPVNTAEL	210648	387950	0.17	-1.8
107976	*Bre*	ATQVYPKL	626010	343031	0.21	1.8
224171	*C330027C09Rik*	AQNDIEHLF	296259	355661	0.61	-1.2

26885	*Casp8ap2*	VTVLNVDHL	112568	295624	0.06	-2.6
75565	*Ccdc101*	ELLTELHQL	109977	377688	0.06	-3.4
76041	*Ccdc125*	NAVLNQRYL	53422	80693	0.05	-1.5
77048	*Ccdc41*	AQVENVQRI	77635	214892	0.005	-2.8
12445	*Ccnd3*	AAVIAHDFL	195002	116726	0.31	1.7
12449	*Ccnf*	SQAVNKQQI	46629	75687	0.48	-1.6
12501	*Cd3e*	YSGLNQRAV	92673	193576	0.05	-2.1
26364	*Cd97*	KLLSNINSVF	177758	812560	0.03	-4.6
12576	*Cdkn1b*	FGPVNHEEL	2765989	8474581	0.003	-3.1
216859	*Centb1*	INQIYEARV	242126	288600	0.58	-1.2
12649	*Chek1*	TGPSNVDKL	5073167	2489307	0.02	2.0
243764	*Chrm2*	TVIGYWPL	57092	112673	0.17	-2.0
70349	*Copb1*	IALRYVAL	579237	197990	0.01	2.9
67876	*Coq10b*	ISFEFRSL	46870	41415	0.74	1.1
71745	*Cul2*	VINSFVHV	163398	55024	0.06	3.0
12767	*Cxcr4*	VVFQFQHI	165482	63047	0.003	2.6
52635	*D12Ertd551e*	RQLENGTTL	405701	580254	0.28	-1.4
218978	*D14Ertd436e*	SQHVNLDQL	274085	752598	0.04	-2.7
69601	*Dab2ip*	VSPTNPTKL	148633	180896	0.52	-1.2
67755	*Ddx47*	KTFLFSATM	76238	60088	0.50	1.3
13207	*Ddx5*	NQAINPKLLQL	20171	115009	0.02	-5.7
13209	*Ddx6*	INFDFPKL	867416	2016311	0.02	-2.3
13204	*Dhx15*	TLLNVYHAF	72069	827769	0.01	-11.5
13433	*Dnmt1*	LSLENGTHTL	1108884	124482	0.02	8.9
94176	*Dock2*	SMVQNRVFL	182029	261777	0.56	-1.4
233115	*Dpy19l3*	SVVAFHNL	114097	81273	0.53	1.4
13424	*Dync1h1*	ASYEFVQRL	2301414	1013239	0.08	2.3
13669	*Eif3s10*	QSIEFSRL	3422637	1081266	0.02	3.2
73830	*Eif3s12*	VGITYQHI	1223275	1317577	0.88	-1.1
54709	*Eif3s2*	FGPINSVAF	63523	12000	0.03	5.3
10436	*Emg1*	VSVEYTEKM	82838	67514	0.30	1.2
13819	*Epas1*	SNYLFTKL	704947	1387353	0.05	-2.0

66366	*Ergic3*	QQLDVEHNL	193740	378788	0.07	-2.0
212377	*F730047E07Rik*	VGVTYRTL	152126	70387	0.12	2.2
107035	*Fbxo38*	SAFSFRTL	270808	192573	0.34	1.4
14137	*Fdft1*	ISLEFRNL	198436	214993	0.81	-1.1
57778	*Fmnl1*	LQYEFTHL	1190978	933913	0.45	1.3
22379	*Fmnl3*	LQYEFTKL	1171553	1937711	0.05	-1.7
108655	*Foxp1*	QQLQQQHLL	518392	2217859	0.01	-4.3
14359	*Fxr1h*	EIVTFERL	336814	132963	0.08	2.5
14790	*Grcc10*	SAPENAVRM	3734208	2535151	0.66	1.5
68153	*Gtf2e2*	SGYKFGVL	972201	381917	0.06	2.5
69237	*Gtpbp4*	QILSDFPKL	1361649	526685	0.03	2.6
15016	*H2-Q5*	AMAPRTLLL	2156626	5661232	0.06	-2.6
15019	*H2-Q8*	FAYEGRDYI	2248785	1616977	0.29	1.4
15201	*Hells*	KVLVFSQM	263746	52922	0.06	5.0
15257	*Hipk1*	SLLTNHVTL	504676	412123	0.63	1.2
319189	*Hist2h2bb*	SVYVYKVL	5932574	2470423	0.02	2.4
98758	*Hnrpf*	FSPLNPVRV	101312	82387	0.49	1.2
16145	*Igtp*	IVAENTKTSL	12000	185606	0.001	-15.5
23918	*Impdh2*	GGIQNVGHI	29380844	17244971	0.01	1.7
16329	*Inpp1*	YMGTNIHSL	37800	259193	0.09	-6.9
76582	*Ipo11*	TIILFTKV	99616	76723	0.41	1.3
338523	*Jhdm1d*	SSIQNGKYTL	182192	886026	0.02	-4.9
110308	*Krt5*	AAYMNKVEL	12000	1280013	0.04	-106.7
110310	*Krt7*	AAYTNKVEL	12000	1458071	0.00	-121.5
72416	*Lrpprc*	AAIENIEHL	159927	163843	0.95	-1.0
26413	*Mapk1*	VGPRYTNL	15419971	25982309	0.02	-1.7
212307	*Mapre2*	QLNEQVHSL	130503	225894	0.28	-1.7
17168	*Mare*	SAVKNLQQL	2084131	2315130	0.64	-1.1
13728	*Mark2*	ASIQNGKDSL	358351	132344	0.03	2.7
217615	*Mgea6*	RNQVYTQL	179246	186263	0.89	-1.0
67014	*Mina*	VVYIYHSL	86825	230698	0.10	-2.7
17427	*Mns1*	KIIEFANI	12000	149382	0.01	-12.4

219135	*Mtmr6*	VGIENIHVM	46307	499479	0.06	-10.8
17886	*Myh9*	RVAEFTTNL	469290	706492	0.28	-1.5
67938	*Mylc2b*	SLGKNPTDAYL	4571127	1100520	0.01	4.2
67608	*Narf*	AAYGFRNI	141414	67434	0.11	2.1
67563	*Narfl*	VAYGFRNI	104022	12000	0.01	8.7
17992	*Ndufa4*	VNVDYSKL	6565000	10705177	0.06	-1.6
227197	*Ndufs1*	AKLVNQEVL	502906	132505	0.07	3.8
18034	*Nfkb2*	AALQNLEQL	275269	482669	0.53	-1.8
64652	*Nisch*	TNQDFIQRL	395004	137997	0.12	2.9
97895	*Nlrp4f*	TPVKNIDTV	44098	40329	0.72	1.1
68979	*Nol11*	TQLIQTHVL	344889	793732	0.04	-2.3
211548	*Nomo1*	SSLVNKEDV	45468	76970	0.07	-1.7
66394	*Nosip*	TSVRFTQL	3233929	3880050	0.30	-1.2
18145	*Npc1*	VAVVNKVDI	337841	629999	0.13	-1.9
101706	*Numa1*	VSILNRQVL	678568	917089	0.33	-1.4
70699	*Nup205*	VNNEFEKL	109609	505624	0.01	-4.6
18813	*Pa2g4*	AQFKFTVL	55142	12000	0.02	4.6
66923	*Pbrm1*	SQVYNDAHI	70006	242803	0.02	-3.5
18538	*Pcna*	RAEDNADTLAL	49024	73938	0.24	-1.5
56426	*Pdcd10*	ILQTFKTVA	71543	309466	0.001	-4.3
207728	*Pde2a*	IKNENQEVI	308207	19230	0.02	16.0
23986	*Peci*	SSYTFPKM	274504	246774	0.82	1.1
18626	*Per1*	YTLRNQDTF	59957	12000	0.04	5.0
56612	*Pfdn5*	SMYVPGKL	199322	12000	0.04	16.6
73739	*Pgea1*	ASLSNLHSL	284222	107706	0.13	2.6
233489	*Picalm*	VAFDFTKV	1026391	970576	0.78	1.1
83490	*Pik3ap1*	YGLKNLTAL	74055	1614342	0.001	-21.8
231329	*Polr2b*	IGPTYYQRL	152006	456565	0.06	-3.0
66385	*Ppp1r7*	RAIENIDTL	1371985	2449039	0.06	-1.8
73699	*Ppp2r1b*	VADKFSEL	123409	43924	0.10	2.8
214572	*Prmt7*	VIVEFRDL	127033	144217	0.65	-1.1
16913	*Psmb8*	GGVVNMYHM	312549	367796	0.77	-1.2

15170	*Ptpn6*	AQYKFIYV	2532589	774524	0.03	3.3
108911	*Rcc2*	AAYRNLGQNL	323760	12000	0.04	27.0
19696	*Rel*	TTLIFQKL	52627	69335	0.46	-1.3
26374	*Rfwd2*	YVVDNIDHL	103204	182323	0.17	-1.8
74734	*Rhoh*	YSVANHNSFL	257883	790683	0.02	-3.1
68477	*Rmnd5a*	WAVSNREML	12000	74507	0.02	-6.2
20054	*Rps15*	VGVYNGKTF	94776	119204	0.32	-1.3
267019	*Rps15a*	VIVRFLTV	561265	225633	0.15	2.5
267019	*Rps15a*	ISPRFDVQL	487134	141860	0.08	3.4
20115	*Rps7*	VNFEFPEF	315064	460959	0.42	-1.5
20133	*Rrm1*	YGIRNSLLI	2099554	1995074	0.76	1.1
20133	*Rrm1*	FQIVNPHLL	8238219	6013528	0.18	1.4
235623	*Scap*	SGYDFSRL	215152	296647	0.35	-1.4
13722	*Scye1*	SGLVNHVPL	121736	137771	0.81	-1.1
69608	*Sec24d*	ESQSVIHNL	154217	185515	0.71	-1.2
20338	*Sell1*	QALKYFNL	342436	120117	0.06	2.9
20393	*Sgk*	STLTYSRM	301908	67793	0.02	9.7
432572	*Specc1*	TSLAFESRL	84991	12000	0.005	7.1
20740	*Spna2*	KALINADEL	4367025	7898847	0.05	-1.8
20815	*Srpk1*	ISGVNGTHI	1498558	1233136	0.46	1.2
20443	*St3gal4*	GAVKNLTYF	84768	217591	0.12	-2.6
20848	*Stat3*	ATLVFHNL	8702796	6298873	0.13	1.4
71728	*Stk11ip*	SALRFLNL	64125	250911	0.01	-3.9
319565	*Syne2*	DLIQTIHEL	64700	102995	0.25	-1.6
108143	*Taf9*	SGLKYVNV	1888511	964010	0.04	2.0
216198	*Tcp11l2*	KATTNIVEM	39337	92718	0.19	-2.4
217353	*Tmc6*	ASYLFRGL	42599	104982	0.01	-2.5
67511	*Tmed9*	VIGNYRTQL	12000	82097	0.01	-6.8
67878	*Tmem33*	QSIAFISRL	312693	175900	0.23	1.8
21916	*Tmod1*	SSIVNKEGL	676206	64142	0.004	10.5
320938	*Tnpo3*	GLLEIAHSL	12000	40153	0.10	-3.3
22057	*Tob1*	ISYLYNKL	137897	149667	0.87	-1.1

21973	*Top2a*	NSMVLFDHV	187186	12000	0.02	15.6
21990	*Tph1*	WGTIFREL	57225	54992	0.90	1.0
21990	*Tph1*	FSLENEVGGL	614142	228500	0.07	2.7
12751	*Tpp1*	IQRVNTEFM	56506	48388	0.66	1.2
100683	*Trrap*	QDPVFQKL	63687	47544	0.41	1.3
56085	*Ubqln1*	RALSNLESI	402950	278602	0.31	1.4
224826	*Ubr2*	VAYKFPELL	36262	33865	0.62	1.1
216825	*Usp22*	FAVVNHQGTL	264021	179805	0.52	1.5
230895	*Vps13d*	INIHYTQL	81820	133304	0.41	-1.6
80743	*Vps16*	VSFTYRYL	1178420	1298769	0.70	-1.1
77573	*Vps33a*	GSLANHTSI	235063	137938	0.23	1.7
233405	*Vps33b*	ASLVNADKL	1141391	750579	0.41	1.5
65114	*Vps35*	FSEENHEPL	100228	72330	0.36	1.4
22375	*Wars*	YTVENAKDII	164659	97161	0.18	1.7
69544	*Wdr5b*	AALENDKTI	995387	846431	0.64	1.2
74781	*Wipi2*	SGYKFFSL	108304	60742	0.28	1.8
22446	*Xlr3b*	VAAANREVL	1019171	12000	0.02	84.9
68090	*Yif1a*	SGYKYVGM	172942	68389	0.14	2.5
66980	*Zdhhc6*	SVIKFENL	57125	76621	0.59	-1.3

Table AII.VI: List of genes coding for MHC I peptides differentially expressed in neoplastic (EL4) versus normal thymocytes and involved in carcinogenesis.

Functional Classification	Gene Symbol	Gene ID	Sequence	p-values	EL4/Thy
Transcription	*Pa2g4* [1-4]	18813	AQFKFTVL	0.018	4.6
	Top2a [5]	21973	NSMVLFDHV	0.020	15.6
	Dnmt1 [6, 7]	13433	LSLENGTHTL	0.022	8.9
	Pfdn5 [8, 9]	56612	SMYVPGKL	0.036	16.6
	Per1 [10, 11]	18626	YTLRNQDTF	0.036	5.0
	Foxp1 [12, 13]	108655	QQLQQQHLL	0.010	-4.3
	Bach2 [14, 15]	12014	EQLEFIHDI	0.015	-20.2
	Ddx5 [16, 17]	13207	NQAINPKLLQL	0.021	-5.7
Cell Differentiation	*Ptpn6* [18, 19]	15170	AQYKFIYV	0.028	3.3
	RhoH [20, 21]	74734	YSVANHNSFL	0.019	-3.1
	Pi3kap1 [22, 23]	83490	YSVANHNSFL	0.001	-21.8
Cell Cycle	*Cdkn1b* [24-26]	12576	FGPVNHEEL	0.003	-3.1
Apoptosis	*Sgk* [27-29]	20393	STLTYSRM	0.022	9.7
	Pdcd10 [30]	56426	ILQTFKTVA	0.001	-4.3
Signal Transduction	*Cxcr4* [31, 32]	12767	VVFQFQHI	0.003	2.6
	CD97 [33, 34]	26364	KLLSNINSVF	0.033	-4.6
Translation	*Eif3s10* [35, 36]	13669	QSIEFSRL	0.015	3.2
	Eif3s2 [35, 36]	54709	FGPINSVAF	0.035	5.3
Miscellaneous	*Dhx15* [37]	13204	TLLNVYHAF	0.017	-11.5
	Gtpbp4 [38]	69237	QILSDFPKL	0.034	2.6
	Igtp [39]	16145	IVAENTKTSL	0.001	-15.5

References

1. Squatrito, M., Mancino, M., Donzelli, M., Areces, L.B., and Draetta, G.F., *EBP1 is a nucleolar growth-regulating protein that is part of pre-ribosomal ribonucleoprotein complexes.* Oncogene, 2004. **23**(25): p. 4454-65.

2. Zhang, Y., Wang, X.W., Jelovac, D., Nakanishi, T., Yu, M.H., Akinmade, D., Goloubeva, O., Ross, D.D., Brodie, A., and Hamburger, A.W., *The ErbB3-binding protein Ebp1 suppresses androgen receptor-mediated gene transcription and tumorigenesis of prostate cancer cells.* Proc Natl Acad Sci U S A, 2005. **102**(28): p. 9890-5.

3. Ahn, J.Y., Liu, X., Liu, Z., Pereira, L., Cheng, D., Peng, J., Wade, P.A., Hamburger, A.W., and Ye, K., *Nuclear Akt associates with PKC-phosphorylated Ebp1, preventing DNA fragmentation by inhibition of caspase-activated DNase.* EMBO J, 2006. **25**(10): p. 2083-95.

4. Liu, Z., Ahn, J.Y., Liu, X., and Ye, K., *Ebp1 isoforms distinctively regulate cell survival and differentiation.* Proc Natl Acad Sci U S A, 2006. **103**(29): p. 10917-22.

5. Murphy, A.J., Hughes, C.A., Barrett, C., Magee, H., Loftus, B., O'Leary, J.J., and Sheils, O., *Low-level TOP2A amplification in prostate cancer is associated with HER2 duplication, androgen resistance, and decreased survival.* Cancer Res, 2007. **67**(6): p. 2893-8.

6. Brown, K.D. and Robertson, K.D., *DNMT1 knockout delivers a strong blow to genome stability and cell viability.* Nat Genet, 2007. **39**(3): p. 289-90.

7. Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., Baylin, S.B., and Vogelstein, B., *DNMT1 and DNMT3b cooperate to silence genes in human cancer cells.* Nature, 2002. **416**(6880): p. 552-6.

8. Fujioka, Y., Taira, T., Maeda, Y., Tanaka, S., Nishihara, H., Iguchi-Ariga, S.M., Nagashima, K., and Ariga, H., *MM-1, a c-Myc-binding protein, is a candidate for a tumor suppressor in leukemia/lymphoma and tongue cancer.* J Biol Chem, 2001. **276**(48): p. 45137-44.

9. Hoffman, B., Amanullah, A., Shafarenko, M., and Liebermann, D.A., *The proto-oncogene c-myc in hematopoietic development and leukemogenesis.* Oncogene, 2002. **21**(21): p. 3414-21.

10. Chen, S.T., Choo, K.B., Hou, M.F., Yeh, K.T., Kuo, S.J., and Chang, J.G., *Deregulated expression of the PER1, PER2 and PER3 genes in breast cancers.* Carcinogenesis, 2005. **26**(7): p. 1241-6.

11. Gery, S., Komatsu, N., Kawamata, N., Miller, C.W., Desmond, J., Virk, R.K., Marchevsky, A., McKenna, R., Taguchi, H., and Koeffler, H.P., *Epigenetic silencing of the candidate tumor suppressor gene Per1 in non-small cell lung cancer.* Clin Cancer Res, 2007. **13**(5): p. 1399-404.

12. Banham, A.H., Beasley, N., Campo, E., Fernandez, P.L., Fidler, C., Gatter, K., Jones, M., Mason, D.Y., Prime, J.E., Trougouboff, P., Wood, K., and Cordell, J.L., *The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p.* Cancer Res, 2001. **61**(24): p. 8820-9.

13. Wlodarska, I., Veyt, E., De Paepe, P., Vandenberghe, P., Nooijen, P., Theate, I., Michaux, L., Sagaert, X., Marynen, P., Hagemeijer, A., and De Wolf-Peeters, C., *FOXP1, a gene highly expressed in a subset of diffuse large B-cell lymphoma, is recurrently targeted by genomic aberrations.* Leukemia, 2005. **19**(8): p. 1299-305.

14. Gutierrez, N.C., Ocio, E.M., de Las Rivas, J., Maiso, P., Delgado, M., Ferminan, E., Arcos, M.J., Sanchez, M.L., Hernandez, J.M., and San Miguel, J.F., *Gene expression profiling of B lymphocytes and plasma cells from Waldenstrom's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals.* Leukemia, 2007. **21**(3): p. 541-9.

15. Yoshida, C., Yoshida, F., Sears, D.E., Hart, S.M., Ikebe, D., Muto, A., Basu, S., Igarashi, K., and Melo, J.V., *Bcr-Abl signaling through the PI-3/S6 kinase pathway inhibits nuclear translocation of the transcription factor Bach2, which represses the antiapoptotic factor heme oxygenase-1.* Blood, 2007. **109**(3): p. 1211-9.

16. Causevic, M., Hislop, R.G., Kernohan, N.M., Carey, F.A., Kay, R.A., Steele, R.J., and Fuller-Pace, F.V., *Overexpression and poly-ubiquitylation of the DEAD-box RNA helicase p68 in colorectal tumours.* Oncogene, 2001. **20**(53): p. 7734-43.

17. Yang, L., Lin, C., and Liu, Z.R., *Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation.* Mol Cancer Res, 2005. **3**(6): p. 355-63.

18. Cuevas, B., Lu, Y., Watt, S., Kumar, R., Zhang, J., Siminovitch, K.A., and Mills, G.B., *SHP-1 regulates Lck-induced phosphatidylinositol 3-kinase phosphorylation and activity.* J Biol Chem, 1999. **274**(39): p. 27583-9.

19. Zhang, Q., Wang, H.Y., Marzec, M., Raghunath, P.N., Nagasawa, T., and Wasik, M.A., *STAT3- and DNA methyltransferase 1-mediated epigenetic silencing of SHP-1 tyrosine phosphatase tumor suppressor gene in malignant T lymphocytes.* Proc Natl Acad Sci U S A, 2005. **102**(19): p. 6948-53.

20. Duquette, M.L., Huber, M.D., and Maizels, N., *G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas.* Cancer Res, 2007. **67**(6): p. 2586-94.

21. Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R., *Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas.* Nature, 2001. **412**(6844): p. 341-6.

22. Okada, T., Maeda, A., Iwamatsu, A., Gotoh, K., and Kurosaki, T., *BCAP: the tyrosine kinase substrate that connects B cell receptor to phosphoinositide 3-kinase activation.* Immunity, 2000. **13**(6): p. 817-27.

23. Yamazaki, T., Takeda, K., Gotoh, K., Takeshima, H., Akira, S., and Kurosaki, T., *Essential immunoregulatory role for BCAP in B cell development and function.* J Exp Med, 2002. **195**(5): p. 535-45.

24. Chien, W.M., Rabin, S., Macias, E., Miliani de Marval, P.L., Garrison, K., Orthel, J., Rodriguez-Puebla, M., and Fero, M.L., *Genetic mosaics reveal both cell-autonomous and cell-nonautonomous function of murine p27Kip1.* Proc Natl Acad Sci U S A, 2006. **103**(11): p. 4122-7.

25. Fu, Y., Fang, Z., Liang, Y., Zhu, X., Prins, P., Li, Z., Wang, L., Sun, L., Jin, J., Yang, Y., and Zha, X., *Overexpression of integrin beta1 inhibits proliferation of hepatocellular carcinoma cell SMMC-7721 through preventing Skp2-dependent degradation of p27 via PI3K pathway.* J Cell Biochem, 2007. **102**(3): p. 704-18.

26. Nakayama, K.I. and Nakayama, K., *Ubiquitin ligases: cell-cycle control and cancer.* Nat Rev Cancer, 2006. **6**(5): p. 369-81.

27. Buse, P., Tran, S.H., Luther, E., Phu, P.T., Aponte, G.W., and Firestone, G.L., *Cell cycle and hormonal control of nuclear-cytoplasmic localization of the serum- and glucocorticoid-inducible protein kinase, Sgk, in mammary tumor cells. A novel convergence point of anti-proliferative and proliferative cell signaling pathways.* J Biol Chem, 1999. **274**(11): p. 7253-63.

28. Sakoda, H., Gotoh, Y., Katagiri, H., Kurokawa, M., Ono, H., Onishi, Y., Anai, M., Ogihara, T., Fujishiro, M., Fukushima, Y., Abe, M., Shojima, N., Kikuchi, M., Oka, Y., Hirai, H., and Asano, T., *Differing roles of Akt and serum- and glucocorticoid-regulated kinase in glucose metabolism, DNA synthesis, and oncogenic activity.* J Biol Chem, 2003. **278**(28): p. 25802-7.

29. Tessier, M. and Woodgett, J.R., *Serum and glucocorticoid-regulated protein kinases: variations on a theme.* J Cell Biochem, 2006. **98**(6): p. 1391-407.

30. Chen, P.Y., Chang, W.S., Chou, R.H., Lai, Y.K., Lin, S.C., Chi, C.Y., and Wu, C.W., *Two non-homologous brain diseases-related genes, SERPINI1 and PDCD10, are tightly linked by an asymmetric bidirectional promoter in an evolutionarily conserved manner.* BMC Mol Biol, 2007. **8**: p. 2.

31. Spoo, A.C., Lubbert, M., Wierda, W.G., and Burger, J.A., *CXCR4 is a prognostic marker in acute myelogenous leukemia.* Blood, 2007. **109**(2): p. 786-91.

32. Tavor, S., Petit, I., Porozov, S., Avigdor, A., Dar, A., Leider-Trejo, L., Shemtov, N., Deutsch, V., Naparstek, E., Nagler, A., and Lapidot, T., *CXCR4 regulates migration and development of human acute myelogenous leukemia stem cells in transplanted NOD/SCID mice.* Cancer Res, 2004. **64**(8): p. 2817-24.

33. Aust, G., Eichler, W., Laue, S., Lehmann, I., Heldin, N.E., Lotz, O., Scherbaum, W.A., Dralle, H., and Hoang-Vu, C., *CD97: a dedifferentiation marker in human thyroid carcinomas.* Cancer Res, 1997. **57**(9): p. 1798-806.

34. Wobus, M., Huber, O., Hamann, J., and Aust, G., *CD97 overexpression in tumor cells at the invasion front in colorectal cancer (CC) is independently regulated of the canonical Wnt pathway.* Mol Carcinog, 2006. **45**(11): p. 881-6.

35. Dong, Z. and Zhang, J.T., *Initiation factor eIF3 and regulation of mRNA translation, cell growth, and cancer.* Crit Rev Oncol Hematol, 2006. **59**(3): p. 169-80.

36. Zhang, L., Pan, X., and Hershey, J.W., *Individual overexpression of five subunits of human translation initiation factor eIF3 promotes malignant transformation of immortal fibroblast cells.* J Biol Chem, 2007. **282**(8): p. 5790-800.

37. Nakagawa, Y., Yoshida, A., Numoto, K., Kunisada, T., Wai, D., Ohata, N., Takeda, K., Kawai, A., and Ozaki, T., *Chromosomal imbalances in malignant peripheral nerve sheath tumor detected by metaphase and microarray comparative genomic hybridization.* Oncol Rep, 2006. **15**(2): p. 297-303.

38. Lee, H., Kim, D., Dan, H.C., Wu, E.L., Gritsko, T.M., Cao, C., Nicosia, S.V., Golemis, E.A., Liu, W., Coppola, D., Brem, S.S., Testa, J.R., and Cheng, J.Q., *Identification and characterization of putative tumor suppressor NGB, a GTP-binding protein that interacts with the neurofibromatosis 2 protein.* Mol Cell Biol, 2007. **27**(6): p. 2103-19.

39. Zhang, H.M., Yuan, J., Cheung, P., Luo, H., Yanagawa, B., Chau, D., Stephan-Tozy, N., Wong, B.W., Zhang, J., Wilson, J.E., McManus, B.M., and Yang, D., *Overexpression of interferon-gamma-inducible GTPase inhibits coxsackievirus B3-induced apoptosis through the activation of the phosphatidylinositol 3-kinase/Akt pathway and inhibition of viral replication.* J Biol Chem, 2003. **278**(35): p. 33011-9.

Table AII.VII: List of primers for quantitative real-time PCR analyses.

Gene symbol	Ref Seq	Gene ID	Primer_A	Primer_B	Probe
2900073G15RIK	NM_026064	67268	cgcttgctttgatgaggaa	gtcagcagctccctcaggt	Universal Library probe: #79
9230105E10RIK	NM_175677	319236	ggccacaaaacagctctcat	cagcctttgcagaactacctg	Universal Library probe: #68
Atxn10	NM_016843.2	54138	gagcctcacaccgaggataa	tggacgtcatttcacacaaga	Universal Library probe: #92
Bach2	NM_007521	12014	catctcttcctctgcccagt	agacatgccgttcaaaccat	Universal Library probe: #89
Ccnd3	NM_007632.2	12445	gcatactggatgctggaggt	ggtagcgatccaggtagttca	Universal Library probe: #88
Ccnf	NM_007634.2	12449	cggaaacacagggactacg	tccacctgcacaggcttt	Universal Library probe: #3
Cd97	NM_011925	26364	atccagccacggtcaactac	ttgttgggtttgagtctccat	Universal Library probe: #11
Cdkn1b	NM_009875	12576	gagcagtgtccagggatgag	tctgttggccctttgtttt	Universal Library probe: #62
Chek1	NM_007691.2	12649	gagggaaggccatatccagt	ttgttcaggcatccctatgtc	Universal Library probe: #25
Copb1	NM_033370.3	70349	ccgcttggtggaattaaaag	ggatgtccatgaccagatcc	Universal Library probe: #15
Cxcr4	NM_009911.3	12767	tggaaccgatcagtgtgagt	gggcaggaagatcctattga	Universal Library probe: #38
Ddx47	NM_026360	67755	tgggaaggccattacttttg	agcttctttccaatcaagtgct	Universal Library probe: #11
Ddx5	NM_007840	13207	agaggtgatgggcctatttg	caagcgacaagctcgaca	Universal Library probe: #66
Ddx6	NM_007841.3	13209	caaaaagtgcactgcctcaa	ggagttgcagaaaatgatgga	Universal Library probe: #80
Dhx15	NM_007839.2	13204	tcggacagtgatccaaatacat	caggcctcatcaatttcctc	Universal Library probe: #12
Dnmt1	NM_010066.3	13433	ctcattggcttctccactgc	gcccaaatattggctcatactc	Universal Library probe: #107
Eif3s10	NM_010123	13669	gattcagaaggcacctggag	catcccgagtttctcgtctc	Universal Library probe: #93
Eif3s12	NM_028659	73830	gctcaagggtatcgacaggt	tttggcttgtgtctccacat	Universal Library probe: #1
Eif3s2	NM_018799	54709	gaagccatggatgtgacca	tcaaaagccaaatggaagaac	Universal Library probe: #22
Gtpbp4	NM_027000.2	69237	tgggagatgattacattttgga	gccttcccaaatttctggtat	Universal Library probe: #95
Igtp	NM_018738	16145	cggagagctgtggagagaga	tgccatgtttatgaaaagtgtaaaa	Universal Library probe: #108
Krt5	NM_027011.2	110308	ccttcgaaacaccaagcac	gttctggaggttggcacact	Universal Library probe: #110
Krt7	NM_033073	110310	ggagatggccaaccacag	ggcctggagtgtctcaaactt	Universal Library probe: #41
Mapk1	NM_011949.3	26413	gacagagtacgtagccacacgtt	agcccacagaccaaatatcaa	Universal Library probe: #50
Mark2	NM_007928.2	13728	gaaagggacacggagcag	agcagaggtggctgagttg	Universal Library probe: #3
Mns1	NM_008613	17427	gaggcagagacccatcctg	gctcaaagctctcttttggtc	Universal Library probe: #63
Mylc2b	NM_023402	67938	aggaggacctgcacgaca	gtccaggtaggcatcagtgg	Universal Library probe: #60
Narf1	NM_026238	67563	gcagttacttccaagtgaaccaa	gcagtcattcagggagacct	Universal Library probe: #78

Pa2g4	NM_011119	18813	ggtcgtgaccaagtataagatgg	cagacacacctgagctggaa	Universal Library probe: #22
Pcna	NM_011045.2	18538	ctagccatgggcgtgaac	gaatactagtgctaaggtgtctgcat	Universal Library probe: #41
Pdcd10	NM_019745.2	56426	tggcagctgatgacgtagaa	tgccttttcatttaggtcctg	Universal Library probe: #21
Pde2a	NM_001008548	207728	tcggacctactgggaaagc	caggggtatgaccagcactt	Universal Library probe: #64
Per1	NM_011065	18626	tggacttgacacctcttctgtg	tgctttagatcggcagtggt	Universal Library probe: #71
Pfdn5	NM_027044	56612	aggagaagcatgccatgaag	agtgcactctcaggctttgac	Universal Library probe: #100
Pik3ap1	NM_031376	83490	cacggtctctgtgaaagcac	ccaggtccccagaatagacc	Universal Library probe: #91
Psmb8	NM_010724	16913	gttggccaaggagtgcag	cagcatcatgttggaaagca	Universal Library probe: #89
Ptpn6	NM_013545.2	15170	tctggatgccacagtcaatg	ggccagtatgggacacattt	Universal Library probe: #51
Rcc2	NM_173867	108911	ctatcagttggggtccatcg	agaagatgccatccagagtctt	Universal Library probe: #79
Rmnd5a	NM_024288	68477	ttcttccgacaaggaatgct	ttgggtctacagacagaccaga	Universal Library probe: #73
Specc1	NM_001029936.2	432572	aaatgccaggctgcagaa	ctctggctcttcattcacagg	Universal Library probe: #33
Stat3	NM_213659.2	20848	agtctcgcctcctccagac	gctgcttctctgtcactacgg	Universal Library probe: #26
Stk11ip	NM_027886.2	71728	ggtgccccagtttttctacc	aacagggacacaggacaggt	Universal Library probe: #41
Tmed9	NM_026211	67511	taaaagaccccgaggacaag	tgtgggaggtgaaggtgaac	Universal Library probe: #21
Tmod1	NM_021883	21916	ctggtcgagctgaaaatcg	tctccatttccactttgttgc	Universal Library probe: #21
Top2a	NM_011623	21973	aatggaaatatgctgccagtg	cctgttcatgtatcaccttcagtg	Universal Library probe: #48
Tpp1	NM_009906.2	12751	cggatcctagctctcctcaa	gtcaggggtgatggttgaa	Universal Library probe: #88
Ubr2	NM_146078.2	224826	cagcagctctgaagttgtgg	tgtaagttctggcgatcacatc	Universal Library probe: #101

Annexe III : Informations supplémentaires pour le chapitre 5

(The mTOR signaling pathway and its influence on the MHC class I peptide repertoire)

Table AIII.I: List of MHC I-associated peptides eluted from EL4 cells.

Peptide	Gene ID	Gene Symbol	Restriction size	WT/B2M**
AAFVFRKL	75627	*Snapc1*	Kb 8-mers	4.6
AAGINRDSL	234309	*Cbr4*	Db 9-mers	21.4
AAIANHQVL	69192	*Dhx16*	Db 9-mers	74.6
AAIENIEHL	72416	*Lrpprc*	Db 9-mers	100.5
AAIRNYGIEL	17769	*Mthfr*	Db 10-mers	74.4
AAITNKYQL	106298	*Rrn3*	Db 9-mers	488.3
AALDFKNV	70099	*Smc4*	Kb 8-mers	35.1
AALENDKTI	140858	*Wdr5*	Db 9-mers	671.7
AALENDKTI	69544	*Wdr5b*		
AALENTHLL	68523	*1110019N10Rik*	Db 9-mers	77.0
AALIYGKL	66861	*Dnajc10*	Kb 8-mers	8.9
AALKNLPLI	74205	*Acsl3*	Db 9-mers	8.3
AALQNLVKI	16211	*Kpnb1*	Db 9-mers	58.6
AAMKNVTEL	22628	*Ywhag*	Db 9-mers	375.5
AAMLNYTHI	223774	*Alg12*	Db 9-mers	13.1
AAPRNSPTGL	231128	*BC037112*	Db 10-mers	13.1
AAPTNRQIEIL	59015	*Nup160*	Db 11-mers	101.8
AAVANQHSSFV	16551	*Kif11*	Db 11-mers	21.5
AAVGNHVAKL	234664	*Nae1*	Db 10-mers	5.4
AAVKFHNL	14571	*Gpd2*	Kb 8-mers	18.6
AAVLNPRFL	74741	*5730419I09Rik*	Db 9-mers	195.2
AAYGFRNI	67608	*Narf*	Kb 8-mers	30.4
AAYRNLGQNL	108911	*Rcc2*	Db 10-mers	39.9
AGLQNAGRSPTNL	17276	*Mela*	Db 13-mers	12.1
AGVRNPQQHL	18458	*Pabpc1*	Db 10-mers	222.4
AGVVNKYEV	13360	*Dhcr7*	Db 9-mers	92.1
AIIAFKTL	73341	*Arhgef6*	Kb 8-mers	5.6
AIVSFAHV	17685	*Msh2*	Kb 8-mers	15.4
AIVTNVDHF	18292	*Sebox*	Db 9-mers	41.7

AKLVNQEVL	227197	*Ndufs1*	Db 9-mers	290.5
ALVRFVNL	76130	*Las1l*	Kb 8-mers	40.7
AMAPRTLLL	14964	*H2-D1*	Qa1 9-mers	2809.9
AMAPRTLLL	15007	*H2-Q10*		
AMAPRTLLL	15016	*H2-Q5*		
AMYIFLHTV	66844	*Ormdl2*	Kb 9-mers	13.3
ANLLYKYL	70885	*Ints10*	Kb 8-mers	4.6
ANVDFSHL	54711	*Plagl2*	Kb 8-mers	9.7
AQFKFTVL	18813	*Pa2g4*	Kb 8-mers	10.6
AQIRFSNI	12464	*Cct4*	Kb 8-mers	7.1
AQIRNLTVL	100910	*2010209O12Rik*	Db 9-mers	95.3
AQIVNKHLI	108067	*Eif2b3*	Db 9-mers	1361.9
AQLLHVHEI	68770	*Phtf2*	Db 9-mers	31.6
AQMQNHSLEM	21335	*Tacc3*	Db 10-mers	23.7
AQNDIEHLF	224171	*C330027C09Rik*	Qa2 9-mers	267.8
AQQVNRTTL	21343	*Taf6*	Db 9-mers	12.0
AQVRNQGYL	228730	*Gm114*	Db 9-mers	29.5
AQYGNILKHVM	69482	*Nup35*	Db 11-mers	410.4
AQYKFIYV	15170	*Ptpn6*	Kb 8-mers	1451.0
AQYRFIYM	19247	*Ptpn11*	Kb 8-mers	23.5
ASIIFEHL	100177	*Zmym6*	Kb 8-mers	3.3
ASITFEHM	68135	*Eif3h*	Kb 8-mers	16.1
ASIVNKDGL	19015	*Ppard*	Db 9-mers	37.0
ASLQNAEKTM	74841	*Usp38*	Db 10-mers	8.0
ASLRYLGL	223664	*Lrrc14*	Kb 8-mers	7.1
ASLSNLHSL	73739	*Cby1*	Db 9-mers	6.7
ASLVNADKL	233405	*Vps33b*	Db 9-mers	529.7
ASPEFTKL	72722	*2810405J04Rik*	Kb 8-mers	217.7
ASPIFTHV	80744	*BC003993*	Kb 8-mers	194.2
ASVINGHTL	329154	*Ankrd44*	Db 9-mers	20.7
ASVLNKWQM	104721	*Ddx1*	Db 9-mers	19.6
ASVLNVNHI	81702	*Ankrd17*	Db 9-mers	1901.6
ASVQNEAKL	23790	*Coro1c*	Db 9-mers	16.9
ASYEFVQRL	13424	*Dync1h1*	Kb 9-mers	1632.8

ASYLFRGL	217353	*Tmc6*	Kb 8-mers	167.1
ATKYFTNRL	63959	*Slc29a1*	Kb 9-mers	4.5
ATLVFHNL	20848	*Stat3*	Kb 8-mers	1605.7
ATQVYPKL	107976	*Bre*	Kb 8-mers	221.6
AVFAFRAV	67249	*Tbc1d19*	Kb 8-mers	6.2
AVIHFAGL	74246	*Gale*	Kb 8-mers	10.6
AVIIFQHL	N.A.	*N.A.*	Kb 8-mers	4.6
AVIKFLEL	17685	*Msh2*	Kb 8-mers	8.2
AVLRYTKL	19718	*Rfc2*	Kb 8-mers	255.4
AVLSFHNI	27405	*Abcg3*	Kb 8-mers	6.8
AVLSFSTRL	16430	*Stt3a*	Kb 9-mers	14.6
AVVAFVMKM	14972	*H2-K1*	Kb 9-mers	17.5
AVVEFSRNV	51797	*Ctps*	Kb 9-mers	56.1
AVVSFKEL	224860	*Plcl2*	Kb 8-mers	12.0
AVYSYKRL	232339	*Ankrd26*	Kb 8-mers	9.6
DAVKNGDYI	75339	*Mphosph8*	Db 9-mers	74.2
DGVANVEHI	107569	*Nt5c3*	Db 9-mers	4.7
EFIRNQEQM	19179	*Psmc1*	Db 9-mers	5.0
EIIKDIHNL	226519	*Lamc1*	Qa2 9-mers	61.1
EIISFQHL	241296	*Lrrc8a*	Kb 8-mers	62.3
EIISFQHL	433926	*Lrrc8b*		
EIISFQHL	231549	*Lrrc8d*		
EIVTFERL	14359	*Fxr1h*	Kb 8-mers	30.6
ELIRVVHQL	15978	*Ifng*	Qa2 9-mers	27.7
ELLTELHQL	75565	*Ccdc101*	Qa2 9-mers	21.5
EQLFQEHRL	244962	*Snx14*	Qa2 9-mers	47.7
EQNESAHTL	229841	*Cenpe*	Qa2 9-mers	20.9
EQQPQQHNL	23856	*Dido1*	Qa2 9-mers	26.5
EQQWVSHTF	68795	*Zfp650*	Qa2 9-mers	4.1
EQVALVHRL	12181	*Bop1*	Qa2 9-mers	419.7
ESPSYRTL	228829	*Phf20*	Kb 8-mers	8.1
ESYSFEARM	67187	*Zmynd19*	Kb 9-mers	25.2
ETIMNQEKL	218490	*Btf3*	Db 9-mers	24.6
FAHTNIESL	14827	*Pdia3*	Db 9-mers	221.4

FAIRNQIKL	106894	*A630042L21Rik*	Db 9-mers	10.4
FALANHLIKV	13660	*Ehd1*	Db 10-mers	8.2
FALANHLIKV	57440	*Ehd3*		
FATNNSEHI	71702	*Cdc5l*	Db 9-mers	14.8
FAVVNHQGTL	216825	*Usp22*	Db 10-mers	317.1
FAYRFSNL	110816	*Pwp2*	Kb 8-mers	188.6
FGGSNVHVI	14104	*Fasn*	Db 9-mers	5.6
FGIHNGVETL	74182	*Prei4*	Db 10-mers	180.2
FGPVNHEEL	12576	*Cdkn1b*	Db 9-mers	2064.7
FLVQNIHTL	67832	*Bxdc2*	Db 9-mers	11.8
FMATNPEHL	66967	*Edem3*	Db 9-mers	22.6
FMKQNLDEL	108099	*Prkag2*	Db 9-mers	35.5
FQAINAGHI	66860	*Tanc1*	Db 9-mers	20.0
FQIVNPHLL	20133	*Rrm1*	Db 9-mers	389.6
FSEENHEPL	65114	*Vps35*	Db 9-mers	84.8
FSPLNPVRV	98758	*Hnrpf*	Db 9-mers	22.8
FSPLNPVRV	59013	*Hnrph1*		
FSQENTEKI	76131	*Depdc1a*	Db 9-mers	12.4
FVHQNVGEI	434223	*Gm1966*	Db 9-mers	48.1
FVISNYREQL	67392	*4833420G17Rik*	Db 10-mers	24.2
GAIVNGKVL	56784	*Garnl1*	Db 9-mers	25.5
GALENAKAEI	67463	*1200014M14Rik*	Db 10-mers	97.8
GANVNHTTV	14155	*Fem1b*	Db 9-mers	38.4
GAVKNLTYF	20443	*St3gal4*	Db 9-mers	38.1
GAVTNVKVI	15568	*Elavl1*	Db 9-mers	72.3
GAVTNVKVI	15569	*Elavl2*		
GAVTNVKVI	15571	*Elavl3*		
GGIQNVGHI	23918	*Impdh2*	Db 9-mers	129.2
GGVVNMYHM	16913	*Psmb8*	Db 9-mers	1078.1
GIMRFVNI	210544	*Wdr67*	Kb 8-mers	16.7
GQIQNKEVL	107182	*Btaf1*	Db 9-mers	6.6
GQMVNNPKIHL	110750	*Cse1l*	Db 11-mers	4.7
GSLANHTSI	77573	*Vps33a*	Db 9-mers	34.9
GSLKNVTTL	59079	*Erbb2ip*	Db 9-mers	85.8

GSVANKFYV	27041	*G3bp1*	Db 9-mers	26.9
GVFSFSRL	69719	*Cad*	Kb 8-mers	4.6
GVFSFSRL	16428	*Itk*		
GVFSFSRL	57028	*Pdxp*		
GVLKFARL	14790	*Grcc10*	Kb 8-mers	28.0
GVLRFVNL	380959	*Alg10b*	Kb 8-mers	55.5
HAIENIDTF	12190	*Brca2*	Db 9-mers	155.4
HQLQERHQL	20868	*Stk10*	Qa2 9-mers	18.1
HTVQNADQV	21355	*Tap2*	Db 9-mers	7.2
IALRYVAL	70349	*Copb1*	Kb 8-mers	39.4
IAVSFREL	68979	*Nol11*	Kb 8-mers	60.0
IFYYVQKL	207304	*Hectd1*	Kb 8-mers	23.3
IGLAYVNHL	218035	*Vps41*	Kb 9-mers	20.8
IGLTFIHM	98488	*Gtf3c3*	Kb 8-mers	5.5
IGLTYKKL	21821	*Ift88*	Kb 8-mers	4.3
IGPKNYEFL	244962	*Snx14*	Db 9-mers	37.4
IGPRYSSV	320487	*Heatr5a*	Kb 8-mers	4.6
IGPTYYQRL	231329	*Polr2b*	Kb 9-mers	230.0
IGVIFTHV	14000	*Rnasen*	Kb 8-mers	9.2
IGYAFVHL	230721	*Pabpc4*	Kb 8-mers	9.9
IILKYIGM	54208	*Arl6ip1*	Kb 8-mers	36.7
IITGFRNV	12236	*Bub1b*	Kb 8-mers	240.6
IIYNPKNL	75062	*Sf3a3*	Kb 8-mers	15.2
ILMEHIHKL	19921	*Rpl19*	Qa2 9-mers	232.9
ILSVFPKV	241035	*Pkhd1*	Kb 8-mers	35.7
IMVRNIDLV	20955	*Vamp7*	Db 9-mers	35.9
INFDFPKL	13209	*Ddx6*	Kb 8-mers	70.9
INIHYTQL	230895	*Vps13d*	Kb 8-mers	241.4
INLSFNKL	67144	*Lrrc40*	Kb 8-mers	9.7
INQIYEARV	216859	*Centb1*	Kb 9-mers	299.4
IQLMNTAHL	74114	*Crot*	Db 9-mers	28.1
IQRVNTEFM	12751	*Tpp1*	Db 9-mers	11.3
IQVRNPVAL	72154	*Zfp157*	Db 9-mers	88.6
IQYAYTGRL	71778	*Klhl5*	Kb 9-mers	9.7

IREEYPDRI	545486	*Tubb1*	Kb 9-mers	66.0
IREEYPDRI	22151	*Tubb2a*		
IREEYPDRI	73710	*Tubb2b*		
IREEYPDRI	227613	*Tubb2c*		
IREEYPDRI	22154	*Tubb5*		
IREEYPDRI	67951	*Tubb6*		
ISFEFRSL	210582	*Coq10a*	Kb 8-mers	6.0
ISFEFRSL	67876	*Coq10b*		
ISFKFDHL	12340	*Capza1*	Kb 8-mers	905.3
ISGVNGTHI	20815	*Srpk1*	Db 9-mers	192.8
ISILYHQL	68295	*0610011L14Rik*	Kb 8-mers	103.1
ISLDYQHL	72503	*2610507B11Rik*	Kb 8-mers	5.1
ISLEFRNL	14137	*Fdft1*	Kb 8-mers	264.2
ISYLYNKL	22057	*Tob1*	Kb 8-mers	7.7
ISYLYNKL	57259	*Tob2*		
IVAENTKTSL	16145	*Igtp*	Db 10-mers	5.0
IVELFRNL	20848	*Stat3*	Kb 8-mers	6.4
KAFDYPSRL	212281	*A530054K11Rik*	Kb 9-mers	11.2
KAFDYPSRL	431706	*Zfp457*		
KAFDYPSRL	238690	*Zfp458*		
KAFDYPSRL	218314	*Zfp595*		
KAFTYINL	22042	*Tfrc*	Kb 8-mers	208.8
KAIETVHNL	57314	*Th1l*	Qa2 9-mers	183.1
KAILNGIDSI	13006	*Smc3*	Db 10-mers	46.2
KAIVNVIGM	81898	*Sf3b1*	Db 9-mers	28.5
KALINADEL	20740	*Spna2*	Db 9-mers	1472.3
KALSYASL*	78757	*4921505C17Rik*	Kb 8-mers	9.7
KAPDNRETL	68292	*Stt3b*	Db 9-mers	50.4
KAPTFEVQM	73828	*Wdr21*	Kb 9-mers	8.4
KAPVNTAEL	66165	*Bccip*	Db 9-mers	103.7
KAVDFDGRL	207165	*Bptf*	Kb 9-mers	56.2
KAVENSSTAI	19167	*Psma3*	Db 10-mers	124.8
KAVENYLIQM	56330	*Pdcd5*	Db 10-mers	8.2
KAVTNEQEL	545725	*Mterf*	Db 9-mers	8.5

KAYIFEGAL	67048	*2610030H06Rik*	Kb 9-mers	13.0
KFLEQVHQL	20846	*Stat1*	Qa2 9-mers	35.8
KGIGNKTEI	234135	*Whsc1l1*	Db 9-mers	328.9
KGYLFNTV	12530	*Cdc25a*	Kb 8-mers	91.8
KIFEFKETL	227541	*Camk1d*	Kb 9-mers	28.4
KIFTASNV	74383	*Ubap2l*	Kb 8-mers	3.5
KIFVFQNQV	234740	*4932417I16Rik*	Kb 9-mers	5.5
KIIEFANI	17427	*Mns1*	Kb 8-mers	25.9
KILANADTM	20336	*Exoc4*	Db 9-mers	4.4
KILTFDQL	19899	*Rpl18*	Kb 8-mers	10.6
KIQSFINRM	231874	*AU022870*	Kb 9-mers	137.6
KIYYFAAV	104831	*Ptpn23*	Kb 8-mers	11.6
KLHDVEHVL	56612	*Pfdn5*	Qa2 9-mers	91.9
KLLDFGSL	20068	*Rps17*	Kb 8-mers	5.8
KMIINEELM	56347	*Eif3c*	Db 9-mers	6.8
KMLEVFHAI	70349	*Copb1*	Qa2 9-mers	11.5
KNFAFTLV	320571	*4930417M19Rik*	Kb 8-mers	7.7
KNFAFTLV	54670	*Atp8b1*		
KNFAFTLV	241633	*Atp8b4*		
KNIFYKAI	13121	*Cyp51*	Kb 8-mers	5.9
KNLLNVDKI	210293	*Dock10*	Db 9-mers	4.1
KNLRYQLL	17184	*Matr3*	Kb 8-mers	172.0
KNLVNKEVM	14751	*Gpi1*	Db 9-mers	290.7
KNVVYERV	69693	*Pof1b*	Kb 8-mers	268.8
KQLEVVHTL	67490	*1810074P20Rik*	Qa2 9-mers	174.0
KQLVNKEHL	51869	*Rif1*	Db 9-mers	190.6
KQVNNLTNL	244879	*Npat*	Db 9-mers	146.6
KSFDYGNL	15273	*Hivep2*	Kb 8-mers	11.7
KSFLFSAL	170439	*Elovl6*	Kb 8-mers	80.5
KSIAFQNV	235682	*Zfp445*	Kb 8-mers	3.5
KSISNPPGSNL	232341	*Wnk1*	Db 11-mers	15.5
KSISNTSKL	226747	*Ahctf1*	Db 9-mers	3.3
KSLKNYITI	71941	*Cars2*	Db 9-mers	7.9
KSLQYLNL	16977	*Lrrc23*	Kb 8-mers	133.6

KSVINTTLV	12492	*Scarb2*	Db 9-mers	26.9
KSYNFHTGL	74133	*1200011M11Rik*	Db 9-mers	19.2
KSYSFIARM	27967	*Cherp*	Kb 9-mers	8.5
KSYVNPTEL	228850	*B230339M05Rik*	Db 9-mers	14.7
KTFLFSATM	67755	*Ddx47*	Kb 9-mers	21.2
KTFSFKSL	22640	*Zfp1*	Kb 8-mers	61.9
KTIYNVEHL	22644	*Rnf103*	Db 9-mers	60.5
KTLLNPEYL	72238	*Tbc1d5*	Db 9-mers	6.3
KTVVNKDVF	19942	*Rpl27*	Db 9-mers	830.5
KTWRFSNM	218952	*Fermt2*	Kb 8-mers	42.7
KTWRFSNM	108101	*Fermt3*		
KTYQFLNDI	224823	*BC011248*	Kb 9-mers	8.6
KVAEFNNV	17758	*Mtap4*	Kb 8-mers	35.7
KVLDVLHSL	77877	*6030458C11Rik*	Qa2 9-mers	44.2
KVLHFFNV	69608	*Sec24d*	Kb 8-mers	24.7
KVLHFYNV	218811	*Sec24c*	Kb 8-mers	36.0
KVLNNMEIGTSL	18655	*Pgk1*	Db 12-mers	4.0
KVLRFLNV	210544	*Wdr67*	Kb 8-mers	26.0
KVLVFSQM	15201	*Hells*	Kb 8-mers	184.9
KVTTFAQL	83922	*Tsga14*	Kb 8-mers	89.1
KVVEFSEL	433864	*Gm1040*	Kb 8-mers	54.0
KVVKFSYM	20747	*Spop*	Kb 8-mers	17.8
KVVKFSYM	76857	*Spopl*		
KVYNYNHL	19942	*Rpl27*	Kb 8-mers	420.4
KVYTFNSV	67972	*Atp2b1*	Kb 8-mers	15.0
KVYTFNSV	11941	*Atp2b2*		
KVYTFNSV	320707	*Atp2b3*		
KVYTFNSV	381290	*Atp2b4*		
LAPQNKPEL	66540	*3110001A13Rik*	Db 9-mers	29.7
LGVTNFVHM	239554	*Foxred2*	Db 9-mers	12.1
LGYTEKDL	13849	*Ephx1*	Kb 8-mers	26.2
LNYKFPGL	106042	*Prickle1*	Kb 8-mers	284.4
LQYEFTHL	57778	*Fmnl1*	Kb 8-mers	27.5
LQYEFTHL	57778	*Fmnl1*	Kb 8-mers	474.0

LQYEFTKL	71409	*Fmnl2*		
LQYEFTKL	22379	*Fmnl3*		
LSILNSNEHL	70012	*Ccdc21*	Db 10-mers	49.3
LSLENGTHTL	13433	*Dnmt1*	Db 10-mers	109.0
LSMRNTSVM	12176	*Bnip3*	Db 9-mers	22.2
LSPINHNTL	78757	*4921505C17Rik*	Db 9-mers	38.8
LSVRNGATL	170826	*Ppargc1b*	Db 9-mers	58.3
LSYSFAHL	195046	*Nlrp1a*	Kb 8-mers	5.2
LSYSFAHL	637515	*Nlrp1b*		
LSYSFAHL	627984	*Nlrp1c*		
LVYQFKEM	13709	*Elf1*	Kb 8-mers	311.5
LVYQFKEM	56501	*Elf4*		
MAHVNGVHL	382051	*4833426J09Rik*	Db 9-mers	198.3
MAYLFRNI	67763	*Prpsap1*	Kb 8-mers	65.5
MSFQFAHL	666173	*Vps13b*	Kb 8-mers	13.1
MSYLFRNI	212627	*Prpsap2*	Kb 8-mers	63.5
NAIKNHWNSTM	17863	*Myb*	Db 11-mers	163.7
NAIKYVNL	22137	*Ttk*	Kb 8-mers	7.8
NAVKNHWNSTI	17865	*Mybl2*	Db 11-mers	45.2
NAYKFPNL	11761	*Aox1*	Kb 8-mers	19.2
NMISVEHHF	65020	*Zfp110*	Qa2 9-mers	25.9
NMISVEHHF	170936	*Zfp369*		
NQFVNKFNVL	26893	*Cops6*	Db 10-mers	34.5
NQLKNTSTI	13209	*Ddx6*	Db 9-mers	30.8
NQVKNIAEL	50505	*Ercc4*	Db 9-mers	12.8
NQYENAEKL	66884	*Appbp2*	Db 9-mers	9.3
NSFRYNGL	19943	*Rpl28*	Kb 8-mers	27.5
NSIRNLDTI	54138	*Atxn10*	Db 9-mers	123.1
NSVVNPNKATI	54644	*Otud5*	Db 11-mers	7.1
NTVTNKVTL	230700	*Foxj3*	Db 9-mers	30.1
NTYKYAKI	99382	*Abtb2*	Kb 8-mers	143.6
QAIHFANL	319934	*Sbf2*	Kb 8-mers	21.6
QAIKNGQAL	11426	*Macf1*	Db 9-mers	106.4
QALKYFNL	20338	*Sel1l*	Kb 8-mers	120.1

QAPQNKITV	16828	*Ldha*	Db 9-mers	209.2
QGFQLTHSL	22151	*Tubb2a*	Qa2 9-mers	19.3
QGFQLTHSL	73710	*Tubb2b*		
QGFQLTHSL	227613	*Tubb2c*		
QGFQLTHSL	22152	*Tubb3*		
QGFQLTHSL	22153	*Tubb4*		
QGFQLTHSL	22154	*Tubb5*		
QGFQLTHSL	67951	*Tubb6*		
QGPEYIERL	74016	*Phf19*	Kb 9-mers	17.7
QIIPFKTL	66409	*Rsl1d1*	Kb 8-mers	190.5
QLLTVKHEL	73086	*Rps6ka5*	Qa2 9-mers	12.5
QLNEQVHSL	212307	*Mapre2*	Qa2 9-mers	96.5
QLNVVIHQL	73162	*Otud3*	Qa2 9-mers	20.8
QLSDTLHSL	214901	*Chtf18*	Qa2 9-mers	24.7
QQFIYEKL	76582	*Ipo11*	Kb 8-mers	8.4
QQIAFKNL*	319491	*1110029I05Rik*	Kb 8-mers	7.9
QQIAFKNL*	21781	*Tfdp1*		
QQIAFKNL*	211586	*Tfdp2*		
QQLDVEHNL	66366	*Ergic3*	Qa2 9-mers	25.7
QQLQIIHRV	116748	*Lsm10*	Qa2 9-mers	74.7
QQQQQLHSL	192191	*Med9*	Qa2 9-mers	8.9
QSIAFISRL	67878	*Tmem33*	Db 9-mers	302.2
QSIEFSRL	13669	*Eif3s10*	Kb 8-mers	1697.8
QSVAFTKL	240641	*Mphosph1*	Kb 8-mers	25.1
QTVENVEHL	75425	*2610036D13Rik*	Db 9-mers	302.7
QVVQFNRL	56692	*Map2k1ip1*	Kb 8-mers	155.6
QVYTFTERM	20448	*St6galnac4*	Kb 9-mers	18.0
RAAANSETI	76022	*5830417I10Rik*	Db 9-mers	5.0
RAFEFTYV	76355	*Tgds*	Kb 8-mers	27.4
RAIAFQHL	18432	*Mybbp1a*	Kb 8-mers	81.5
RAIENIDTL	66385	*Ppp1r7*	Db 9-mers	18.4
RAIKNDSVV	12468	*Cct7*	Db 9-mers	647.7
RAIPNNQVL	59022	*Edf1*	Db 9-mers	22.0
RALENPDTASL	104831	*Ptpn23*	Db 11-mers	8.7

RALSNLESI	56085	*Ubqln1*	Db 9-mers	90.9
RALSNLESI	244178	*Ubqln3*		
RANQNFDEI	77480	*C330002I19Rik*	Db 9-mers	225.1
RAPAFHQL	268721	*2310021P13Rik*	Kb 8-mers	14.1
RAPQFINL	213109	*Phf3*	Kb 8-mers	59.4
RAVDNQVYV	52633	*Nit2*	Db 9-mers	182.8
RAVKNQIAL	56321	*Aatf*	Db 9-mers	80.9
RAYLFAHV	104662	*Tsr1*	Kb 8-mers	416.7
RAYLFNSV	106369	*Ypel1*	Kb 8-mers	18.1
RAYLFNSV	77864	*Ypel2*		
RAYLFNSV	66090	*Ypel3*		
RAYLFNSV	241525	*Ypel4*		
RLITNSEEI	19299	*Abcd3*	Db 9-mers	75.5
RLLQEAHDL	66884	*Appbp2*	Qa2 9-mers	121.2
RNFIFSRL	15201	*Hells*	Kb 8-mers	640.3
RNLDYARL	20425	*Shmt1*	Kb 8-mers	241.9
RNYEYLIRL	66855	*Tcf25*	Kb 9-mers	38.8
RNYQFDFL	67465	*Sf3a1*	Kb 8-mers	49.5
RQAENGMYI	69719	*Cad*	Db 9-mers	39.1
RQATNQIVM	19155	*Npepps*	Db 9-mers	80.0
RQIEYFRSL	74392	*Specc1l*	Kb 9-mers	5.4
RQILNADAM	66185	*1110037F02Rik*	Db 9-mers	97.7
RQLENGTTL	52635	*D12Ertd551e*	Db 9-mers	511.7
RQVQNTAITL	74522	*Morc2a*	Db 10-mers	13.7
RQVQNTAITL	240069	*Morc2b*		
RSIQNAQFL	11432	*Acp2*	Db 9-mers	197.0
RSISFSNM	18201	*Nsmaf*	Kb 8-mers	332.0
RSPENPPSKEL	66570	*Cenpm*	Db 11-mers	114.3
RTYSFLNL	230770	*Tmem39b*	Kb 8-mers	81.6
RTYTYEKL	12387	*Ctnnb1*	Kb 8-mers	334.4
RVAEFTTNL	17886	*Myh9*	Kb -9mers	75.9
RVFQNEVLGTL	224045	*Eif2b5*	Db 11-mers	5.3
RVLIFSQM	12648	*Chd1*	Kb 8-mers	66.4
RVLIFSQM	244059	*Chd2*		

RVLIFSQM	216848	*Chd3*		
RVLIFSQM	107932	*Chd4*		
RVLIFSQM	269610	*Chd5*		
RVLIFSQM	320790	*Chd7*		
RVLIFSQM	93761	*Smarca1*		
RVLIFSQM	93762	*Smarca5*		
RVVANSEEI	11666	*Abcd1*	Db 9-mers	6.4
RVYYFNHI	23988	*Pin1*	Kb 8-mers	18.0
SAALNKDFL	74081	*Cep350*	Db 9-mers	12.0
SAARFALL	107723	*Slc12a6*	Kb 8-mers	5.3
SAFEHTQM	12033	*Bcap29*	Kb 8-mers	6.0
SAFSFRTL	107035	*Fbxo38*	Kb 8-mers	157.3
SALANYIHL	58245	*Gpr180*	Db 9-mers	113.7
SALENAENHV	217216	*BC030867*	Db 10-mers	62.5
SALENGRYEL	217578	*Baz1a*	Db 10-mers	11.8
SALQNAESDRL	229841	*Cenpe*	Db 11-mers	497.4
SALRFLNL	71728	*Stk11ip*	Kb 8-mers	27.1
SALRFQAM	74015	*Fcho1*	Kb 8-mers	56.4
SALVFTRL	56700	*0610031J06Rik*	Kb 8-mers	21.2
SAPENAVRM	14790	*Grcc10*	Db 9-mers	333.1
SAPRNFVENF	58523	*Elp2*	Db 10-mers	123.7
SAVENILEHL	121022	*Mrps6*	Db 10-mers	3.1
SAVIFRTL	326622	*Upf2*	Kb 8-mers	145.1
SAVKNDYEM	68275	*Rpa1*	Db 9-mers	703.3
SAVKNLQQL	17168	*Mare*	Db 9-mers	1262.4
SAVTNSGVHL	66067	*Gtpbp8*	Db 10-mers	14.9
SAYEYIHAL	53417	*Hif3a*	Kb 9-mers	9.4
SAYRFSGV	226747	*Ahctf1*	Kb 8-mers	10.0
SGFVFTRL	20585	*Hltf*	Kb 8-mers	74.7
SGFYYKTRI	217431	*Nol10*	Kb 9-mers	7.9
SGIKNPVSV	75812	*Tasp1*	Db 9-mers	29.9
SGLKYVNV	108143	*Taf9*	Kb 8-mers	685.2
SGLTYIKI	12211	*Birc6*	Kb 8-mers	4.1
SGLVNHVPL	13722	*Scye1*	Db 9-mers	95.8

SGPTYIKL	57869	*Adck2*	Kb 8-mers	54.2
SGVDYRGV	78651	*Lsm6*	Kb 8-mers	4.5
SGVRFNNV	53321	*Cntnap1*	Kb 8-mers	5.5
SGVSNPHVI	208146	*Yeats2*	Db 9-mers	111.4
SGYDFSRL	235623	*Scap*	Kb 8-mers	39.9
SGYEFIHKL	70120	*Yars2*	Kb 9-mers	34.1
SGYIYHKL	19704	*Upf1*	Kb 8-mers	795.3
SGYKFFSL	74781	*Wipi2*	Kb 8-mers	102.3
SGYKFGVL	68153	*Gtf2e2*	Kb 8-mers	419.4
SGYKYVGM	68090	*Yif1a*	Kb 8-mers	80.7
SGYSFTHI	67019	*Actr6*	Kb 8-mers	170.6
SIAAFIQRL	224727	*Bat3*	Kb 9-mers	10.2
SIQNFSKL	19645	*Rb1*	Kb 8-mers	7.4
SKYDFPKL	21822	*Tgtp*	Kb 8-mers	30.6
SLGKNPTDAYL	67938	*Mylc2b*	Db 11-mers	31.1
SLITNKVVM	18950	*Pnp*	Kb 9-mers	51.6
SLLTNHVTL	15257	*Hipk1*	Db 9-mers	4.2
SLVKNVAHM	52521	*Zfp622*	Db 9-mers	120.9
SLYKLRTL	14412	*Slc6a13*	Kb 8-mers	18.9
SLYLFNHL	245423	*Gm364*	Kb 8-mers	15.3
SMGKNPTDEYL	67268	*2900073G15Rik*	Db 11-mers	481.1
SMMDVDHQI	12465	*Cct5*	Qa2 9-mers	215.1
SMVQNRVFL	94176	*Dock2*	Db 9-mers	242.6
SMYVPGKL	56612	*Pfdn5*	Kb 8-mers	62.9
SNFHFAVL	666173	*Vps13b*	Kb 8-mers	12.4
SNFTFSHL	68897	*Disp1*	Kb 8-mers	11.6
SNIHYHTL	110749	*Chaf1b*	Kb 8-mers	40.6
SNLKYILV	29808	*Mga*	Kb 8-mers	45.0
SNLKYSLL	70028	*Dopey2*	Kb 8-mers	5.8
SNLRYLNL	241568	*Lrrc4c*	Kb 8-mers	4.0
SNLYYKYL	56697	*Akap10*	Kb 8-mers	3.7
SNPEFRQL	338523	*Jhdm1d*	Kb 8-mers	6.7
SNVKYVML	19206	*Ptch1*	Kb 8-mers	9.8
SNVVFKLL	14976	*H2-Ke2*	Kb 8-mers	9.6

SNYRVSLL	110078	*Pygb*	Kb 8-mers	11.8
SNYRVSLL	110095	*Pygl*		
SNYRVSLL	19309	*Pygm*		
SQAVNKQQI	12449	*Ccnf*	Db 9-mers	38.3
SQFPNAEKM	94176	*Dock2*	Db 9-mers	7.3
SQHVNLDQL	218978	*D14Ertd436e*	Db 9-mers	140.6
SQIKNEINI	230866	*C230096C10Rik*	Db 9-mers	11.9
SQLIDTHLL	226562	*Bat2d*	Qa2 9-mers	31.9
SQLKNADVEL	11772	*Ap2a2*	Db 10-mers	18.5
SQLNELHHL	231123	*BC023882*	Qa2 9-mers	23.7
SQLPVDHIL	12847	*Copa*	Qa2 9-mers	93.3
SQLQAEHTL	232906	*Grlf1*	Qa2 9-mers	20.3
SQLRNADVEL	11771	*Ap2a1*	Db 10-mers	7.5
SQLRNEVAI	18760	*Prkcm*	Db 9-mers	57.7
SQLRNEVAI	75292	*Prkcn*		
SQLRNEVAI	101540	*Prkd2*		
SQMTNLQELHL	231549	*Lrrc8d*	Db 11-mers	6.0
SQVINSHSV	208154	*Btla*	Db 9-mers	57.5
SQVRNNVYM	A2A7B5	*Prdm2*	Db 9-mers	13.7
SQVRNNVYM	14411	*Slc6a12*		
SQVYNDAHI	66923	*Pbrm1*	Db 9-mers	62.9
SRIVFRHL	64659	*Mrps14*	Kb 8-mers	399.9
SRVSFTSL	545216	*EG545216*	Kb 8-mers	26.9
SRVSFTSL	170772	*Glcci1*		
SSFQFHNRM	240066	*BC066107*	Kb 9-mers	25.1
SSIQNGKYTL	338523	*Jhdm1d*	Db 10-mers	303.9
SSIVNKEGL	21916	*Tmod1*	Db 9-mers	37.9
SSIYFKSL	58180	*Hic2*	Kb 8-mers	5.0
SSLINHKSV	22720	*Zfp62*	Db 9-mers	11.3
SSLQNHNHQL	109331	*Rnf20*	Db 10-mers	7.7
SSLSNAKDV	320632	*Ascc3l1*	Db 9-mers	16.9
SSLSNEHVL	237615	*Ankrd52*	Db 9-mers	28.8
SSLVNKEDV	211548	*Nomo1*	Db 9-mers	50.8
SSLVNKEDVL	211548	*Nomo1*	Db 10-mers	18.0

SSPENKNWL	76843	*Dtl*	Db 9-mers	196.1
SSPKFSEI	54636	*Wdr45*	Kb 8-mers	419.3
SSPSNKFFF	235626	*Setd2*	Db 9-mers	64.6
SSSLNQEKI	105203	*BC016423*	Db 9-mers	10.6
SSVKFNPV	223499	*Wdsof1*	Kb 8-mers	28.8
SSVQNYFHL	67095	*Trak1*	Db 9-mers	11.8
SSVRFSYM	230082	*Nol6*	Kb 8-mers	9.3
SSVSNKTTL	227446	*2310035C23Rik*	Db 9-mers	49.8
SSVYFRSV	72368	*2310045N01Rik*	Kb 8-mers	8.8
SSYAYTKV	194388	*D230004J03Rik*	Kb 8-mers	36.1
SSYFFGKL	27405	*Abcg3*	Kb 8-mers	9.6
SSYKFNHL	110606	*Fntb*	Kb 8-mers	32.2
SSYNFIRHM	232853	*5730403M16Rik*	Kb 9-mers	8.7
SSYNYRVV	16438	*Itpr1*	Kb 8-mers	6.7
SSYSFRHL	54397	*Ppt2*	Kb 8-mers	220.4
SSYTFPKM	23986	*Peci*	Kb 8-mers	74.1
SSYTFPKMM	23986	*Peci*	Kb 9-mers	18.5
STFEFHSI	22222	*Ubr1*	Kb 8-mers	74.2
STIRNAQSI	17969	*Ncf1*	Db 9-mers	47.6
STIVYYKL	66637	*5730449L18Rik*	Kb 8-mers	5.8
STLIYRNM	75660	*Lin37*	Kb 8-mers	60.5
STLTYSRM	20393	*Sgk1*	Kb 8-mers	175.8
STVRNADVI	18671	*Abcb1a*	Db 9-mers	735.0
STVRNADVI	18669	*Abcb1b*		
STYKFFEV	66480	*Rpl15*	Kb 8-mers	675.3
SVIKFENL	66980	*Zdhhc6*	Kb 8-mers	495.6
SVMENSKVLGEAM	21894	*Tln1*	Db 13-mers	80.0
SVMENSKVLGESM	70549	*Tln2*	Db 13-mers	5.6
SVVAFHNL	233115	*Dpy19l3*	Kb 8-mers	48.9
SVYVYKVL	319178	*Hist1h2bb*	Kb 8-mers	309.1
SVYVYKVL	319181	*Hist1h2bg*		
SVYVYKVL	319182	*Hist1h2bh*		
SVYVYKVL	319183	*Hist1h2bj*		
SVYVYKVL	319184	*Hist1h2bk*		

SVYVYKVL	319186	*Hist1h2bm*		
SVYVYKVL	319188	*Hist1h2bp*		
SVYVYKVL	319189	*Hist2h2bb*		
SVYVYKVL	665596	*RP23-38E20.1*		
TAIENSWIHL	67493	*Mett10d*	Db 10-mers	5.5
TALRFLEL	75669	*Pik3r4*	Kb 8-mers	4.4
TAPQYYRL	67789	*Dalrd3*	Kb 8-mers	23.0
TAVLNLHEHL	319565	*Syne2*	Db 10-mers	14.2
TAVVNRVFDKL	16912	*Psmb9*	Db 11-mers	13.9
TGIKNGVHFL	50907	*Preb*	Db 10-mers	157.0
TGIRNLEWL	218989	*6720456H20Rik*	Db 9-mers	18.7
TGPSNVDKL	12649	*Chek1*	Db 9-mers	102.7
TGVTNRDLI	71435	*Arhgap21*	Db 9-mers	50.9
TIIIFHSL	14049	*Eya2*	Kb 8-mers	267.9
TIIIFHSL	14050	*Eya3*		
TIILFTKV	76582	*Ipo11*	Kb 8-mers	10.9
TLNDLIHNI	80986	*Ckap2*	Qa2 9-mers	27.1
TMKTNLEYL	67241	*Smc6*	Db 9-mers	76.3
TNLRYLAL	11771	*Ap2a1*	Kb 8-mers	9.0
TNLRYLAL	11772	*Ap2a2*		
TNQDFIQRL	64652	*Nisch*	Kb 9-mers	335.0
TNVLFNHL	71919	*Rpap3*	Kb 8-mers	19.5
TNYKFFML	75965	*Zdhhc20*	Kb 8-mers	10.7
TQLIQTHVL	68979	*Nol11*	Qa2 9-mers	164.4
TQPLNHYFI	18796	*Plcb2*	Db 9-mers	1343.3
TSIAFKNI	226778	*Mark1*	Kb 8-mers	19.6
TSIAFKNI	17169	*Mark3*		
TSIKNQTQL	78697	*Pus7*	Db 9-mers	206.7
TSPINPQHM	217351	*Tnrc6c*	Db 9-mers	17.4
TSVENHEFL	226747	*Ahctf1*	Db 9-mers	1011.6
TSVQFMKL	14104	*Fasn*	Kb 8-mers	10.6
TSVRFTQL	66394	*Nosip*	Kb 8-mers	1435.0
TSVVFNKL	66136	*Znrd1*	Kb 8-mers	5.9
TSYRFLAL	231123	*BC023882*	Kb 8-mers	270.7

TTYKYEMI	223691	*Eif3eip*	Kb 8-mers	26.8
TTYKYFAL	214290	*Zcchc6*	Kb 8-mers	55.5
VAAANREVL	574437	*Xlr3a/b*	Db 9-mers	907.2
VADKFSEL	73699	*Ppp2r1b*	Kb 8-mers	16.4
VAFAYKNV	67437	*Ssr3*	Kb 8-mers	130.7
VAFDFTKV	233489	*Picalm*	Kb 8-mers	257.6
VAHTFVIGV	18813	*Pa2g4*	Kb 9-mers	72.8
VAIGNPVHL	11906	*Zfhx3*	Db 9-mers	11.4
VALAFRHL	75547	*Akap13*	Kb 8-mers	1027.3
VALEFTHL	57913	*Lrdd*	Kb 8-mers	11.6
VAVKNSGGFL	52635	*D12Ertd551e*	Db 10-mers	21.5
VAVSNLEKV	68145	*Etaa1*	Db 9-mers	18.8
VAVTFSERL	26570	*Slc7a11*	Kb 9-mers	22.6
VAVVNKVDI	18145	*Npc1*	Db 9-mers	18.5
VAYGFRNI	67563	*Narfl*	Kb 8-mers	92.1
VEFFLDAL	320631	*Abca15*	Kb 8-mers	209.8
VFYAVKVL	20393	*Sgk1*	Kb 8-mers	16.7
VGFRFPIL	19395	*Rasgrp2*	Kb 8-mers	19.4
VGFTFPNRL	26562	*Ncdn*	Kb 9-mers	34.9
VGIENIHVM	219135	*Mtmr6*	Db 9-mers	116.0
VGLINKDSV	192195	*Ash1l*	Db 9-mers	83.0
VGLLYRLL	269254	*Setx*	Kb 8-mers	3.0
VGLRYYTGV	20585	*Hltf*	Kb 9-mers	17.2
VGPKNKTSI	22224	*Usp10*	Db 9-mers	100.4
VGPRYTNL	26413	*Mapk1*	Kb 8-mers	488.6
VGVTYRTL	212377	*F730047E07Rik*	Kb 8-mers	37.0
VGVYNGKTF	633683	*EG633683*	Db 9-mers	18.0
VGVYNGKTF	20054	*Rps15*		
VGYRFVTAI	64652	*Nisch*	Kb 9-mers	387.6
VGYRYETL	109674	*Ampd2*	Kb 8-mers	249.2
VIHIFSHI	70603	*Mutyh*	Kb 8-mers	9.0
VINAVTHAL	71911	*Bdh1*	Qa2 9-mers	10.2
VINSFVHV	71745	*Cul2*	Kb 8-mers	25.9
VINVFHHL	56706	*Ccnl1*	Kb 8-mers	78.3

VIQDFVKM	106143	*Cggbp1*	Kb 8-mers	46.2
VITEFARI	212528	*Trmt1*	Kb 8-mers	23.2
VITHFVHV	19395	*Rasgrp2*	Kb 8-mers	35.3
VITKFDHL	277939	*C2cd3*	Kb 8-mers	15.2
VITKFINV	240168	*Rasgrp3*	Kb 8-mers	15.6
VIVEFRDL	214572	*Prmt7*	Kb 8-mers	12.9
VIVKFAQL	18572	*Pdcd11*	Kb 8-mers	13.9
VIVRFLTV	267019	*Rps15a*	Kb 8-mers	264.9
VIVRFLTVM	267019	*Rps15a*	Kb 9-mers	23.9
VKYLFTGL	59025	*Usp14*	Kb 8-mers	10.2
VMLENYSHL	408067	*9630025I21Rik*	Db 9-mers	17.6
VMLENYSHL	67911	*Zfp169*		
VMLENYSHL	22693	*Zfp30*		
VMLENYSHL	245368	*Zfp300*		
VMLENYSHL	328274	*Zfp459*		
VMLENYSHL	78251	*Zfp712*		
VMLENYSHL	235907	*Zfp71-rs1*		
VMLENYSHL	212276	*Zfp748*		
VMLENYSHL	233056	*Zfp790*		
VMLENYSHL	238693	*Zfp817*		
VMLENYSHL	170763	*Zfp87*		
VMLENYSHL	22754	*Zfp92*		
VMNYVEHDL	12537	*Cdc2l1*	Qa2 9-mers	9.6
VNDIFERI	319178	*Hist1h2bb*	Kb 8-mers	9.2
VNDIFERI	319181	*Hist1h2bg*		
VNDIFERI	319182	*Hist1h2bh*		
VNDIFERI	319183	*Hist1h2bj*		
VNDIFERI	319184	*Hist1h2bk*		
VNDIFERI	319186	*Hist1h2bm*		
VNDIFERI	319188	*Hist2h2bp*		
VNDIFERI	319189	*Hist1h2bb*		
VNDIFERI	319190	*Hist2h2be*		
VNDIFERI	78303	*Hist3h2ba*		
VNDIFERI	382522	*Hist3h2bb*		

VNFVHTNL	67891	*Rpl4*	Kb 8-mers	1167.5
VNIEFKDL	11307	*Abcg1*	Kb 8-mers	306.4
VNLIYHQRI	212569	*Zfp273*	Kb 9-mers	5.3
VNLIYHQRI	22746	*Zfp85-rs1*		
VNQFFKTML	94176	*Dock2*	Kb 9-mers	9.4
VNQKFNNL	218977	*Dlg7*	Kb 8-mers	648.4
VNRVFDKL	16912	*Psmb9*	Kb 8-mers	531.1
VNSIFQHL	77700	*9130208D14Rik*	Kb 8-mers	3.0
VNSIFQHL	434223	*Gm1966*		
VNSIFQHL	74558	*Gvin1*		
VNSNFYLRM*	68845	*Pih1d1*	Kb 9-mers	9.0
VNTHFSHL*	73668	*Ttc21b*	Kb 8-mers	15.3
VNVDYSKL	17992	*Ndufa4*	Kb 8-mers	752.5
VNVRFTGV	75007	*4930504E06Rik*	Kb 8-mers	16.8
VNVRFTGV	235461	*B230380D07Rik*		
VNYHYMSQV	12211	*Birc6*	Kb 9-mers	5.2
VNYLHRNV	641050	*LOC641050*	Kb 8-mers	223.9
VNYRHLAL	20020	*Polr2a*	Kb 8-mers	133.4
VNYRVPNM	18803	*Plcg1*	Kb 8-mers	248.5
VNYYFERNM	75725	*Phf14*	Kb 9-mers	12.9
VQFLYREL	433956	*Heatr2*	Kb 8-mers	4.3
VQIKNDVFI	76788	*2410127E18Rik*	Db 9-mers	7.6
VQMKFRLL	268697	*Ccnb1*	Kb 8-mers	39.6
VQMKFRLL	434175	*EG434175*		
VQNKNSSYF	22151	*Tubb2a*	Db 9-mers	7.2
VQNKNSSYF	73710	*Tubb2b*		
VQNKNSSYF	227613	*Tubb2c*		
VQNKNSSYF	22154	*Tubb5*		
VQWEYGRL	97541	*Qars*	Kb 8-mers	420.6
VQYKFSHL	330129	*EG330129*	Kb 8-mers	743.3
VQYKFSHL	14534	*Gcn5l2*		
VQYKFSHL	18519	*Pcaf*		
VQYLYRVF	77591	*Ddx10*	Kb 8-mers	11.4
VSFTYRYL	80743	*Vps16*	Kb 8-mers	878.0

VSILNRQVL	101706	*Numa1*	Db 9-mers	284.5
VSIQFYHL	271564	*Vps13a*	Kb 8-mers	132.1
VSISFKSL	14727	*Gp49a*	Kb 8-mers	4.6
VSISFKSL	22201	*Uba1*		
VSLKYAHM	11566	*Adss*	Kb 8-mers	3.8
VSPLFQKL	216443	*Mars*	Kb 8-mers	17.0
VSPQNVHHSYL	100040505	*LOC100040505*	Db 11-mers	96.0
VSPRNSLEVL	242418	*Wdr32*	Db 10-mers	4.1
VSQYYPKL	15441	*Hp1bp3*	Kb 8-mers	24.8
VSTKFEHL	320714	*D030016E14Rik*	Kb 8-mers	57.1
VSVEYTEKM	14791	*Emg1*	Kb 9-mers	336.1
VSVTNEHLM*	22035	*Tnfsf10*	Db 9-mers	9.0
VSYKNPSLM*	114741	*Supt16h*	Db 9-mers	8.9
VSYKYSKV	16004	*Igf2r*	Kb 8-mers	6.9
VSYLFSHV	328099	*AU021838*	Kb 8-mers	70.0
VSYLFSHV	19139	*Prps1*		
VSYLFSHV	75456	*Prps1l1*		
VSYLFSHV	110639	*Prps2*		
VSYNHTNI	382252	*A830080D01Rik*	Kb 8-mers	23.6
VSYRNIEEL	100043998	*LOC100043998*	Db 9-mers	19.1
VSYRNIEEL	108989	*Tpr*		
VTIHYNKL	18693	*Pick1*	Kb 8-mers	380.4
VTLVFEHI	12567	*Cdk4*	Kb 8-mers	5.6
VTLVFEHI	640611	*LOC640611*		
VTVDFSKL	117599	*Helb*	Kb 8-mers	16.0
VTVLNVDHL	26885	*Casp8ap2*	Db 9-mers	40.2
VVFQFQHI	12767	*Cxcr4*	Kb 8-mers	154.4
VVYIYHSL	67014	*Mina*	Kb 8-mers	34.4
WAVSNREML	68477	*Rmnd5a*	Db 9-mers	45.5
WSLANHEYF	140580	*Elmo1*	Db 9-mers	4.7
YAGSNFPEHI	66713	*Actr2*	Db 10-mers	543.5
YAIKNIHGV	13430	*Dnm2*	Db 9-mers	22.2
YALYNNWEHM	545270	*LOC545270*	Db 10-mers	112.1
YAMENTRQTI	20185	*Ncor1*	Db 10-mers	5.6

YAMIYRNL	17246	*Mdm2*	Kb 8-mers	8.5
YGINNIREL	66590	*Farsa*	Db 9-mers	7.8
YGIQNHHEV	320817	*Atad2b*	Db 9-mers	28.6
YGIRNSLLI	20133	*Rrm1*	Db 9-mers	323.2
YGISNEKPEV	54624	*Paf1*	Db 10-mers	10.1
YGLLNVTKI	75415	*Arhgap12*	Db 9-mers	23.2
YGNILKHVM	69482	*Nup35*	Qa2 9-mers	5.6
YGYHFPEL	67134	*Nol5a*	Kb 8-mers	10.0
YGYSNRVVDLM	100039258	*LOC100039258*	Db 11-mers	7.3
YGYSNRVVDLM	100039840	*LOC100039840*		
YGYSNRVVDLM	100040109	*LOC100040109*		
YGYSNRVVDLM	100040634	*LOC100040634*		
YGYSNRVVDLM	100041236	*LOC100041236*		
YGYSNRVVDLM	100041399	*LOC100041399*		
YGYSNRVVDLM	100041831	*LOC100041831*		
YGYSNRVVDLM	100044981	*LOC100044981*		
YGYSNRVVDLM	100047129	*LOC100047129*		
YGYSNRVVDLM	100047352	*LOC100047352*		
YGYSNRVVDLM	100048117	*LOC100048117*		
YGYSNRVVDLM	100048253	*LOC100048253*		
YGYSNRVVDLM	640374	*LOC640374*		
YGYSNRVVDLM	215974	*EG215974*		
YKNVNQEVV	53415	*Htatip2*	Db 9-mers	76.1
YMGTNIHSL	16329	*Inpp1*	Db 9-mers	45.9
YMLANLTHL	66689	*Klhl28*	Db 9-mers	8.9
YQHINSYQL	13990	*Smarcad1*	Db 9-mers	28.7
YQHLNVVEM	214230	*Pak6*	Db 9-mers	16.7
YQKENKDVI	17420	*Mnat1*	Db 9-mers	245.2
YQNQEIHNL	213895	*Bms1*	Qa2 9-mers	269.3
YQPYNKDWI	13877	*Erh*	Db 9-mers	195.4
YQVGNLAGHTL	66610	*Abi3*	Db 11-mers	7.1
YSGLNQRAV	12501	*Cd3e*	Db 9-mers	26.4
YSGTNTGNVHI	72313	*Fryl*	Db 11-mers	7.9
YSVANHNSFL	74734	*Rhoh*	Db 10-mers	619.7

YSYQNRYHF	12914	*Crebbp*	Db 9-mers	5.1
YSYQNRYHF	328572	*Ep300*		
YTIQNRVDI	20842	*Stag1*	Db 9-mers	7.9
YTLLNKAPEYL	23881	*G3bp2*	Db 11-mers	6.3
YTVANKEYV	67458	*Ergic1*	Db 9-mers	52.9
YTVENAKDII	22375	*Wars*	Db 10-mers	36.7
YVLHNSNTM	432448	*EG432448*	Db 9-mers	8.4
YVLHNSNTM	21762	*Psmd2*		
YVVDNIDHL	621205	*EG621205*	Db 9-mers	137.3
YVVDNIDHL	100045235	*LOC100045235*		
YVVDNIDHL	26374	*Rfwd2*		

* Only detected following rapamycin exposure
** Fold change ≥ 3 to < 5 correspond to low abundance MHC I peptides only detected in the WT condition

Table AIII.II: Kinetic profiles of MHC I-associated peptides eluted from control and rapamycin-treated EL4 cells.

unchanged MHC I peptides (across the kinetic profiles)

Peptide	Gene ID	Gene Symbol	Restriction size	WT/B2M*	FC T6/T0	FC T12/T0	FC T24/T0	FC T48/T0	P T6/T0	P T12/T0	P T24/T0	P T48/T0
AAIANHQVL	69192	*Dhx16*	Db 9-mers	74.6	1.02	1.40	-1.24	1.32	0.9608	0.3085	0.6754	0.3386
AAIRNYGIEL	17769	*Mthfr*	Db 10-mers	74.4	-1.14	1.18	1.00	-2.17	0.6477	0.5663	0.9999	0.0534
AALENDKTI	140858	*Wdr5*	Db 9-mers	671.7	1.03	1.25	-1.04	1.21	0.9558	0.5486	0.9140	0.6194
AALENDKTI	69544	*Wdr5b*										
AALENTHLL	68523	*1110019N10Rik*	Db 9-mers	77.0	-1.07	1.08	-1.90	1.52	0.8577	0.7429	0.1422	0.4371
AALQNLVKI	16211	*Kpnb1*	Db 9-mers	58.6	1.09	2.41	2.44	2.18	0.8887	0.1447	0.0685	0.5546
AAPTNRQIEIL	59015	*Nup160*	Db 11-mers	101.8	-1.13	1.26	1.10	-1.34	0.6566	0.3478	0.6722	0.2138
AAVANQHSSFV	16551	*Kif11*	Db 11-mers	21.5	-1.23	1.44	1.13	1.49	0.4910	0.2146	0.6442	0.1628
AAVKFHNL	14571	*Gpd2*	Kb 8-mers	18.6	-1.04	1.72	1.21	1.34	0.8653	0.1885	0.3649	0.0161
AQLLHVHEI	68770	*Phtf2*	Db 9-mers	31.6	-1.39	1.06	-1.03	-1.29	0.1046	0.8652	0.8466	0.2847
AQMQNHSLEM	21335	*Tacc3*	Db 10-mers	23.7	-1.78	1.57	1.48	1.05	0.3774	0.6754	0.0607	0.8940
AQNDIEHLF	224171	*C330027C09Rik*	Qa2 9-mers	267.8	1.02	1.14	1.26	-1.22	0.9722	0.6878	0.5671	0.6973
ASIVNKDGL	19015	*Ppard*	Db 9-mers	37.0	-1.01	1.25	-1.15	-1.34	0.9793	0.3500	0.6504	0.3147
ASPIFTHV	80744	*BC003993*	Kb 8-mers	194.2	1.27	1.67	1.68	1.69	0.5225	0.0853	0.1376	0.4847
ASVLNKWQM	104721	*Ddx1*	Db 9-mers	19.6	-1.12	1.69	1.46	1.31	0.7907	0.1752	0.4717	0.5851
ASVLNVNHI	81702	*Ankrd17*	Db 9-mers	1901.6	-1.08	1.21	1.11	1.22	0.8600	0.5338	0.7701	0.6048
ASVQNEAKL	23790	*Coro1c*	Db 9-mers	16.9	1.53	3.15	1.66	1.96	0.4471	0.0336	0.1115	0.3496
ASYEFVQRL	13424	*Dync1h1*	Kb 9-mers	1632.8	1.07	1.61	1.60	1.34	0.7787	0.1799	0.0064	0.6044
ASYLFRGL	217353	*Tmc6*	Kb 8-mers	167.1	-4.59	-2.94	-1.49	-1.76	0.5095	0.4630	0.7244	0.6335
AVVAFVMKM	14972	*H2-K1*	Kb 9-mers	17.5	1.71	2.28	2.17	1.78	0.1319	0.1749	0.1138	0.0790
EIIKDIHNL	226519	*Lamc1*	Qa2 9-mers	61.1	-1.16	1.17	1.36	-1.10	0.1044	0.5291	0.0817	0.8376
EIISFQHL	241296	*Lrrc8a*	Kb 8-mers	62.3	-1.06	1.30	1.18	-1.15	0.9434	0.7235	0.8019	0.8859
EIISFQHL	433926	*Lrrc8b*										
EIISFQHL	231549	*Lrrc8d*										
ELIRVVHQL	15978	*Ifng*	Qa2 9-mers	27.7	-1.26	-1.12	1.34	-1.75	0.0481	0.6693	0.1897	0.1278
ELLTELHQL	75565	*Ccdc101*	Qa2 9-mers	21.5	1.20	1.45	1.72	1.30	0.6208	0.3991	0.1138	0.7660
EQLFQEHRL	244962	*Snx14*	Qa2 9-mers	47.7	-1.27	1.13	1.32	-1.46	0.4146	0.7605	0.2781	0.4680
EQNESAHTL	229841	*Cenpe*	Qa2 9-mers	20.9	-2.02	1.03	-1.57	-1.88	0.3912	0.9554	0.5828	0.4328
EQQPQQHNL	23856	*Dido1*	Qa2 9-mers	26.5	-1.08	1.16	1.05	1.06	0.8385	0.6944	0.8849	0.8566

EQVALVHRL	12181	*Bop1*	Qa2 9-mers	419.7	-1.17	1.08	1.03	-1.72	0.5626	0.7861	0.8960	0.2380
FAHTNIESL	14827	*Pdia3*	Db 9-mers	221.4	-1.09	-1.11	-1.15	-1.00	0.8418	0.7812	0.7241	0.9992
FGIHNGVETL	74182	*Prei4*	Db 10-mers	180.2	-1.03	1.91	1.38	1.61	0.9214	0.1156	0.2416	0.1449
FMATNPEHL	66967	*Edem3*	Db 9-mers	22.6	-1.02	1.03	18.64	1.23	0.9457	0.9388	0.3555	0.6050
FQIVNPHLL	20133	*Rrm1*	Db 9-mers	389.6	1.07	1.97	2.23	1.33	0.8614	0.2030	0.0241	0.6677
FSEENHEPL	65114	*Vps35*	Db 9-mers	84.8	1.03	1.82	1.25	1.17	0.9146	0.0355	0.4327	0.4517
FSPLNPVRV	98758	*Hnrpf*	Db 9-mers	22.8	-1.19	-1.21	-2.75	-1.14	0.8295	0.7917	0.3585	0.8870
FSPLNPVRV	59013	*Hnrph1*										
FVISNYREQL	67392	*4833420G17Rik*	Db 10-mers	24.2	-1.20	2.31	2.24	1.65	0.7414	0.2137	0.0829	0.2791
GQIQNKEVL	107182	*Btaf1*	Db 9-mers	6.6	1.19	2.16	2.25	1.16	0.7758	0.0808	0.0675	0.8205
GSVANKFYV	27041	*G3bp1*	Db 9-mers	26.9	-1.20	1.41	1.39	1.45	0.7771	0.5116	0.5045	0.5163
IAVSFREL	68979	*Nol11*	Kb 8-mers	60.0	-1.00	1.40	1.92	-1.21	0.9976	0.6233	0.0751	0.6533
IGPKNYEFL	244962	*Snx14*	Db 9-mers	37.4	-1.13	1.30	1.34	1.09	0.5109	0.3597	0.0689	0.7336
IGPTYYQRL	231329	*Polr2b*	Kb 9-mers	230.0	-1.65	-1.25	-1.43	-1.03	0.4823	0.7314	0.5793	0.9636
IQYAYTGRL	71778	*Klhl5*	Kb 9-mers	9.7	1.43	2.38	-1.59	1.56	0.3598	0.0782	0.4464	0.5105
ISFKFDHL	12340	*Capza1*	Kb 8-mers	905.3	-1.47	1.03	2.33	-1.59	0.4745	0.9564	0.1049	0.6215
KAIETVHNL	57314	*Th1l*	Qa2 9-mers	183.1	-1.35	-1.02	-1.11	1.03	0.3925	0.9631	0.7409	0.9185
KAPDNRETL	68292	*Su3b*	Db 9-mers	50.4	-1.72	-1.49	-3.03	-1.25	0.5111	0.6120	0.2948	0.7876
KAVENYLIQM	56330	*Pdcd5*	Db 10-mers	8.2	1.64	2.17	2.29	1.29	0.4256	0.2354	0.0714	0.6643
KFLEQVHQL	20846	*Stat1*	Qa2 9-mers	35.8	-1.21	-1.01	1.71	-1.34	0.3788	0.9710	0.0580	0.6126
KGYLFNTV	12530	*Cdc25a*	Kb 8-mers	91.8	1.70	-1.11	-1.05	1.20	0.4148	0.8836	0.9491	0.7555
KMLEVFHAI	70349	*Copb1*	Qa2 9-mers	11.5	-1.03	1.11	1.50	-1.08	0.9494	0.8425	0.3411	0.9124
KNLRYQLL	17184	*Matr3*	Kb 8-mers	172.0	-1.08	1.53	2.41	1.30	0.8624	0.4471	0.0012	0.7244
KQLEVVHTL	67490	*1810074P20Rik*	Qa2 9-mers	174.0	-1.29	1.04	1.02	-1.02	0.4560	0.9163	0.9608	0.9492
KTLLNPEYL	72238	*Tbc1d5*	Db 9-mers	6.3	1.26	2.55	2.50	2.17	0.4058	0.1708	0.0303	0.2438
KTVVNKDVF	19942	*Rpl27*	Db 9-mers	830.5	-1.26	1.41	-1.04	1.15	0.1821	0.1093	0.7252	0.3240
KTYQFLNDI	224823	*BC011248*	Kb 9-mers	8.6	1.12	1.47	1.35	1.14	0.7394	0.4236	0.4104	0.8509
KVAEFNNV	17758	*Mtap4*	Kb 8-mers	35.7	-2.30	-1.30	-11.03	-2.18	0.5494	0.8516	0.2911	0.5645
KVLDVLHSL	77877	*6030458C11Rik*	Qa2 9-mers	44.2	1.37	-1.07	3.55	1.03	0.5032	0.8011	0.1500	0.9658
KVLHFYNV	218811	*Sec24c*	Kb 8-mers	36.0	1.24	1.45	**3.35**	1.92	0.4471	0.4561	**0.0024**	0.5334
KVTTFAQL	83922	*Tsga14*	Kb 8-mers	89.1	-1.16	1.39	2.03	1.19	0.8509	0.6552	0.2416	0.8457
LGYTEKDL	13849	*Ephx1*	Kb 8-mers	26.2	-4.65	1.84	1.03	-3.18	0.4180	0.5197	0.9749	0.4751
LQYEFTHL	57778	*Fmnl1*	Kb 8-mers	27.5	1.20	2.09	1.68	1.76	0.8369	0.1257	0.2566	0.1973
LSLENGTHTL	13433	*Dnmt1*	Db 10-mers	109.0	1.46	2.48	2.35	1.94	0.3428	0.0776	0.1167	0.1153
LSMRNTSVM	12176	*Bnip3*	Db 9-mers	22.2	-1.16	-1.37	-2.18	-1.43	0.6896	0.3772	0.1144	0.2144

NAVKNHWNSTI	17865	*Mybl2*	Db 11-mers	45.2	-1.35	1.67	1.33	1.72	0.5793	0.3330	0.4647	0.3016
NAYKFPNL	11761	*Aox1*	Kb 8-mers	19.2	-1.17	1.18	1.30	1.11	0.5998	0.6262	0.3992	0.6680
NMISVEHHF	65020	*Zfp110*	Db 9-mers	25.9	-2.08	1.05	-1.12	-1.21	0.2604	0.9163	0.8621	0.7653
NMISVEHHF	170936	*Zfp369*										
NSVVNPNKATI	54644	*Otud5*	Db 11-mers	7.1	1.02	1.83	1.96	1.60	0.9541	0.1871	0.2296	0.0286
QAIHFANL	319934	*Sbf2*	Kb 8-mers	21.6	-1.04	1.42	1.20	1.46	0.8823	0.2496	0.4558	0.2242
QAIKNGQAL	11426	*Macf1*	Db 9-mers	106.4	1.08	1.82	2.00	1.60	0.7363	0.0629	0.0131	0.0227
QAPQNKITV	16828	*Ldha*	Db 9-mers	209.2	-1.15	1.28	-1.03	1.36	0.4311	0.2254	0.8980	0.0509
QLLTVKHEL	73086	*Rps6ka5*	Qa2 9-mers	12.5	-1.00	1.04	1.77	1.13	0.9949	0.9058	0.1259	0.8443
QLNEQVHSL	212307	*Mapre2*	Qa2 9-mers	96.5	-1.06	1.25	1.21	-1.07	0.8899	0.5284	0.6124	0.8551
QLNVVIHQL	73162	*Otud3*	Qa2 9-mers	20.8	-1.13	1.46	1.37	1.10	0.6509	0.3078	0.1832	0.8757
QQFIYEKL	76582	*Ipo11*	Kb 8-mers	8.4	-1.06	1.29	2.05	1.22	0.8014	0.5499	0.0682	0.6174
QQLQIIHRV	116748	*Lsm10*	Qa2 9-mers	74.7	-1.37	1.10	1.12	-1.09	0.3183	0.7397	0.6737	0.8181
QQQQQLHSL	192191	*Med9*	Qa2 9-mers	8.9	-1.17	1.13	1.25	1.04	0.6670	0.7115	0.5179	0.8945
QSIAFISRL	67878	*Tmem33*	Db 9-mers	302.2	-1.06	1.61	1.64	-2.40	0.8886	0.3886	0.2179	0.1991
QSIEFSRL	13669	*Eif3s10*	Kb 8-mers	1697.8	1.05	1.68	1.24	1.09	0.8842	0.0831	0.3549	0.7409
RAIENIDTL	66385	*Ppp1r7*	Db 9-mers	18.4	1.20	1.63	-1.12	1.40	0.7499	0.2932	0.8685	0.6330
RAIKNDSVV	12468	*Cct7*	Db 9-mers	647.7	-1.35	-1.05	-1.93	1.19	0.1920	0.8663	0.1540	0.7083
RALENPDTASL	104831	*Ptpn23*	Db 11-mers	8.7	1.34	2.83	1.93	1.57	0.0424	0.2070	0.0395	0.0398
RLLQEAHDL	66884	*Appbp2*	Qa2 9-mers	121.2	-1.50	-1.13	1.11	-1.16	0.0390	0.7593	0.7468	0.7318
RVYYFNHI	23988	*Pin1*	Kb 8-mers	18.0	1.10	1.35	**3.02**	1.57	0.8402	0.6427	**0.0227**	0.6941
SALANYIHL	58245	*Gpr180*	Db 9-mers	113.7	1.05	1.55	1.62	1.38	0.9335	0.4399	0.2678	0.7689
SALENGRYEL	217578	*Baz1a*	Db 10-mers	11.8	-1.97	1.55	1.17	1.30	0.0469	0.2441	0.5663	0.3085
SALQNAESDRL	229841	*Cenpe*	Db 11-mers	497.4	1.25	2.24	1.82	1.97	0.6720	0.1049	0.2616	0.1388
SAPRNFVENF	58523	*Elp2*	Db 10-mers	123.7	-1.18	1.41	1.24	1.13	0.4531	0.2139	0.0659	0.5032
SAVKNDYEM	68275	*Rpa1*	Db 9-mers	703.3	-1.79	1.08	-1.17	1.08	0.3129	0.8658	0.7875	0.8941
SGIKNPVSV	75812	*Tasp1*	Db 9-mers	29.9	1.10	3.32	1.38	2.46	0.7736	0.1039	0.3312	0.0379
SGLVNHVPL	13722	*Scye1*	Db 9-mers	95.8	-1.14	1.15	-1.10	1.13	0.7447	0.6570	0.7985	0.7155
SGYKFGVL	68153	*Gtf2e2*	Kb 8-mers	419.4	1.03	1.66	1.28	1.58	0.9274	0.1356	0.4522	0.1474
SMMDVDHQI	12465	*Cct5*	Qa2 9-mers	215.1	-1.07	-1.08	-1.01	-1.46	0.8876	0.8884	0.9880	0.5335
SQLIDTHLL	226562	*Bat2d*	Qa2 9-mers	31.9	1.24	1.84	1.93	1.21	0.5105	0.1421	0.0797	0.8025
SQLPVDHIL	12847	*Copa*	Qa2 9-mers	93.3	-1.12	1.21	1.20	-1.30	0.7447	0.4729	0.6012	0.5263
SQLQAEHTL	232906	*Grlf1*	Qa2 9-mers	20.3	-1.01	-1.00	1.21	-1.13	0.9816	0.9950	0.6924	0.8181
SRIVFRHL	64659	*Mrps14*	Kb 8-mers	399.9	1.44	3.23	3.08	1.54	0.7006	0.2438	0.1128	0.7032
SSIVNKEGL	21916	*Tmod1*	Db 9-mers	37.9	-1.08	1.59	-1.44	**-3.70**	0.8275	0.0817	0.3384	**0.0059**

SSLSNEHVL	237615	*Ankrd52*	Db 9-mers	28.8	1.08	1.45	1.42	1.37	0.7063	0.0937	0.3586	0.3230
SSLVNKEDV	211548	*Nomo1*	Db 9-mers	50.8	-1.28	1.06	-1.24	-1.10	0.4487	0.7818	0.4730	0.7809
SSLVNKEDVL	211548	*Nomo1*	Db 10-mers	18.0	1.07	1.74	-1.02	1.01	0.8505	0.1010	0.9626	0.9663
SSPENKNWL	76843	*Dtl*	Db 9-mers	196.1	1.03	2.46	2.01	2.52	0.9330	0.0365	0.0388	0.0212
SSPKFSEI	54636	*Wdr45*	Kb 8-mers	419.3	1.02	1.64	1.16	1.69	0.9390	0.1133	0.5587	0.1452
SSVKFNPV	223499	*Wdsof1*	Kb 8-mers	28.8	-1.38	-1.13	-1.56	1.31	0.4306	0.7188	0.2902	0.5486
SSYNFIRHM	232853	*5730403M16Rik*	Kb 9-mers	8.7	-1.14	1.22	1.32	-1.01	0.7924	0.7067	0.4363	0.9893
STFEFHSI	22222	*Ubr1*	Kb 8-mers	74.2	1.14	1.48	1.27	1.17	0.7248	0.3080	0.5477	0.6910
STVRNADVI	18671	*Abcb1a*	Db 9-mers	735.0	1.05	1.26	-1.06	1.24	0.9220	0.5420	0.8977	0.5885
STVRNADVI	18669	*Abcb1b*										
STYKFFEV	66480	*Rpl15*	Kb 8-mers	675.3	1.22	1.71	1.21	1.48	0.5750	0.0968	0.4436	0.4794
TGIRNLEWL	218989	*6720456H20Rik*	Db 9-mers	18.7	-5.63	3.06	-2.64	-2.37	0.1597	0.3816	0.2479	0.4013
TIIIFHSL	14049	*Eya2*	Kb 8-mers	267.9	1.40	1.68	1.94	1.14	0.5249	0.3180	0.1926	0.8941
TIIIFHSL	14050	*Eya3*										
TLNDLIHNI	80986	*Ckap2*	Qa2 9-mers	27.1	-1.06	1.76	1.67	1.42	0.9187	0.3584	0.2282	0.7226
TNVLFNHL	71919	*Rpap3*	Kb 8-mers	19.5	1.57	2.62	2.21	2.48	0.2169	0.1671	0.1081	0.0919
TQLIQTHVL	68979	*Nol11*	Qa2 9-mers	164.4	-1.09	1.17	1.18	-1.26	0.8049	0.6293	0.6032	0.5531
TQPLNHYFI	18796	*Plcb2*	Db 9-mers	1343.3	1.06	3.68	2.40	-1.21	0.9682	0.1293	0.6141	0.3788
TSVENHEFL	226747	*Ahctf1*	Db 9-mers	1011.6	-1.04	2.05	1.54	2.15	0.9139	0.0406	0.2573	0.0164
VGPKNKTSI	22224	*Usp10*	Db 9-mers	100.4	-1.16	-1.04	-1.44	-1.31	0.7136	0.9255	0.4692	0.6575
VMNYVEHDL	12537	*Cdc2l1*	Qa2 9-mers	9.6	-1.70	-1.55	-1.43	-1.14	0.4411	0.5235	0.5503	0.7839
VNDIFERI	319178	*Hist1h2bb*	Kb 8-mers	9.2	1.70	1.20	2.22	1.45	0.1833	0.6405	0.3101	0.5101
VNDIFERI	319181	*Hist1h2bg*										
VNDIFERI	319182	*Hist1h2bh*										
VNDIFERI	319183	*Hist1h2bj*										
VNDIFERI	319184	*Hist1h2bk*										
VNDIFERI	319186	*Hist1h2bm*										
VNDIFERI	319188	*Hist2h2bp*										
VNDIFERI	319189	*Hist1h2bb*										
VNDIFERI	319190	*Hist2h2be*										
VNDIFERI	78303	*Hist3h2ba*										
VNDIFERI	382522	*Hist3h2bb*										
VNYHYMSQV	12211	*Birc6*	Kb 9-mers	5.2	1.04	1.58	1.51	1.42	0.9196	0.4292	0.2480	0.4730
VQWEYGRL	97541	*Qars*	Kb 8-mers	420.6	1.01	-2.28	-1.02	1.05	0.9723	0.2699	0.8616	0.3312
VSIQFYHL	271564	*Vps13a*	Kb 8-mers	132.1	-1.05	1.43	1.67	1.11	0.9103	0.4145	0.1307	0.9086

VSPLFQKL	216443	*Mars*	Kb 8-mers	17.0	1.17	1.68	1.44	1.74	0.6377	0.2020	0.2052	0.0977
VTVLNVDHL	26885	*Casp8ap2*	Db 9-mers	40.2	-1.05	1.30	1.39	1.27	0.8471	0.1368	0.3556	0.7395
YAGSNFPEHI	66713	*Actr2*	Db 10-mers	543.5	-1.11	1.37	-1.00	-1.01	0.6843	0.2073	0.9871	0.9675
YALYNNWEHM	100041219	*LOC100041219*	Db 10-mers	112.1	1.14	2.15	1.60	1.55	0.8520	0.2935	0.3962	0.6570
YALYNNWEHM	545270	*LOC545270*										
YGIRNSLLI	20133	*Rrm1*	Db 9-mers	323.2	-1.11	1.55	1.71	1.33	0.7460	0.3731	0.0635	0.7293
YGISNEKPEV	54624	*Paf1*	Db 10-mers	10.1	1.00	1.44	1.45	1.04	0.9922	0.3713	0.3356	0.9128
YGLLNVTKI	75415	*Arhgap12*	Db 9-mers	23.2	-1.01	1.72	1.93	1.71	0.9736	0.3114	0.1567	0.4367
YGYSNRVVDLM	100039258	*LOC100039258*	Db 11-mers	7.3	-1.30	1.01	1.35	-1.38	0.6034	0.9918	0.5534	0.5825
YGYSNRVVDLM	100039840	*LOC100039840*										
YGYSNRVVDLM	100040109	*LOC100040109*										
YGYSNRVVDLM	100040634	*LOC100040634*										
YGYSNRVVDLM	100041236	*LOC100041236*										
YGYSNRVVDLM	100041399	*LOC100041399*										
YGYSNRVVDLM	100041831	*LOC100041831*										
YGYSNRVVDLM	100044981	*LOC100044981*										
YGYSNRVVDLM	100047129	*LOC100047129*										
YGYSNRVVDLM	100047352	*LOC100047352*										
YGYSNRVVDLM	100048117	*LOC100048117*										
YGYSNRVVDLM	100048253	*LOC100048253*										
YGYSNRVVDLM	640374	*LOC640374*										
YGYSNRVVDLM	215974	*EG215974*										
YQHINSYQL	13990	*Smarcad1*	Db 9-mers	28.7	1.56	2.45	2.24	1.90	0.0751	0.0830	0.0137	0.0156
YQHLNVVEM	214230	*Pak6*	Db 9-mers	16.7	-1.22	1.04	1.32	-1.12	0.6044	0.9324	0.3837	0.7261
YQNQEIHNL	213895	*Bms1*	Qa2 9-mers	269.3	-1.13	1.13	1.09	-1.06	0.7432	0.7157	0.7950	0.8736
YQPYNKDWI	13877	*Erh*	Db 9-mers	195.4	-1.10	2.87	2.21	2.02	0.8923	0.1811	0.1645	0.3598
YSGTNTGNVHI	72313	*Fryl*	Db 11-mers	7.9	1.03	2.63	2.32	1.51	0.9448	0.3298	0.3101	0.2485
YTVENAKDII	22375	*Wars*	Db 10-mers	36.7	1.38	3.19	2.57	1.57	0.3163	0.2632	0.1183	0.3690
YVVDNIDHL	621205	*EG621205*	Db 9-mers	137.3	1.33	2.99	-1.07	2.67	0.7620	0.2590	0.9229	0.0899
YVVDNIDHL	100045235	*LOC100045235*										
YVVDNIDHL	26374	*Rfwd2*										

overexpressed MHC I peptides (across the kinetic profiles)

Peptide	Gene ID	Gene Symbol	Restriction size	WT/B2M*	FC T6/T0	FC T12/T0	FC T24/T0	FC T48/T0	P T6/T0	P T12/T0	P T24/T0	P T48/T0
AAFVFRKL	75627	*Snapc1*	Kb 8-mers	4.6	-1.25	1.48	1.95	2.01	0.6151	0.3136	0.0991	0.3356
AAGINRDSL	234309	*Cbr4*	Db 9-mers	21.4	1.41	1.38	1.70	1.46	0.6368	0.6430	0.4970	0.5696
AAIENIEHL	72416	*Lrpprc*	Db 9-mers	100.5	-1.17	-1.01	-1.48	1.46	0.7271	0.9719	0.3171	0.4583
AAITNKYQL	106298	*Rrn3*	Db 9-mers	488.3	-1.20	1.27	1.16	1.62	0.5113	0.1784	0.5382	0.0848
AALDFKNV	70099	*Smc4*	Kb 8-mers	35.1	1.68	2.01	3.12	3.23	0.1417	0.2364	0.0822	0.1362
AALIYGKL	66861	*Dnajc10*	Kb 8-mers	8.9	1.12	1.57	2.03	3.20	0.5012	0.0347	0.0026	0.2649
AAMKNVTEL	22628	*Ywhag*	Db 9-mers	375.5	-1.52	-1.15	-1.03	1.49	0.5598	0.7762	0.9603	0.3421
AAMLNYTHI	223774	*Alg12*	Db 9-mers	13.1	1.16	1.08	1.35	2.03	0.6718	0.8310	0.3797	0.3347
AAPRNSPTGL	231128	*BC037112*	Db 10-mers	13.1	-1.05	1.42	1.96	**3.39**	0.8852	0.0905	0.0721	**0.0066**
AAVGNHVAKL	234664	*Nae1*	Db 10-mers	5.4	1.40	2.80	2.42	**4.64**	0.5705	0.1137	0.0479	**0.0003**
AAVLNPRFL	74741	*5730419I09Rik*	Db 9-mers	195.2	1.00	1.58	2.13	2.25	0.9901	0.2377	0.0025	0.0556
AAYGFRNI	67608	*Narf*	Kb 8-mers	30.4	1.36	2.26	2.47	2.95	0.3434	0.0866	0.0144	0.0566
AAYRNLGQNL	108911	*Rcc2*	Db 10-mers	39.9	1.54	2.26	2.88	3.24	0.0650	0.1433	0.0346	0.0733
AGLQNAGRSPTNL	17276	*Mela*	Db 13-mers	12.1	-1.17	-1.03	-1.05	1.18	0.6936	0.9336	0.8854	0.7251
AGVRNPQQHL	18458	*Pabpc1*	Db 10-mers	222.4	1.13	1.68	2.21	**3.41**	0.5989	0.1876	0.0132	**0.0013**
AGVVNKYEV	13360	*Dhcr7*	Db 9-mers	92.1	-1.10	1.36	1.53	**3.81**	0.7488	0.2015	0.3150	**0.0086**
AIVSFAHV	17685	*Msh2*	Kb 8-mers	15.4	-1.05	1.21	1.68	2.93	0.9175	0.6374	0.2695	0.3506
AKLVNQEVL	227197	*Ndufs1*	Db 9-mers	290.5	-1.01	1.86	1.18	2.09	0.9750	0.0580	0.4567	0.0942
ALVRFVNL	76130	*Las1l*	Kb 8-mers	40.7	-1.16	1.51	1.60	3.45	0.8386	0.6624	0.4029	0.4055
AMAPRTLLL	14964	*H2-D1*	Qa1 9-mers	2809.9	1.38	1.41	2.63	**2.92**	0.3505	0.5233	0.0533	**0.0430**
AMAPRTLLL	15007	*H2-Q10*										
AMAPRTLLL	15016	*H2-Q5*										
AMYIFLHTV	66844	*Ormdl2*	Kb 9-mers	13.3	2.49	2.65	**3.81**	2.94	0.2432	0.1892	**0.0062**	0.4439
ANVDFSHL	54711	*Plagl2*	Kb 8-mers	9.7	1.26	1.52	2.75	3.78	0.6298	0.5579	0.0214	0.2079
AQIRNLTVL	100910	*2010209O12Rik*	Db 9-mers	95.3	-1.37	1.45	1.34	1.86	0.1860	0.2645	0.3745	0.1213
AQQVNRTTL	21343	*Taf6*	Db 9-mers	12.0	2.24	2.14	**3.30**	**4.07**	0.0417	0.3884	**0.0319**	**0.0029**
AQYGNILKHVM	69482	*Nup35*	Db 11-mers	410.4	1.19	-1.00	1.50	1.23	0.7515	0.9924	0.3894	0.8513
AQYKFIYV	15170	*Ptpn6*	Kb 8-mers	1451.0	1.14	1.79	1.90	2.14	0.7214	0.1291	0.0714	0.3053
AQYRFIYM	19247	*Ptpn11*	Kb 8-mers	23.5	1.45	1.50	2.79	2.45	0.2726	0.4212	0.0241	0.2335
ASLSNLHSL	73739	*Cby1*	Db 9-mers	6.7	1.57	2.47	2.44	3.18	0.4637	0.1239	0.1288	0.0683
ASLVNADKL	233405	*Vps33b*	Db 9-mers	529.7	-1.07	1.19	1.09	1.94	0.8798	0.5242	0.7785	0.1376
ASPEFTKL	72722	*2810405J04Rik*	Kb 8-mers	217.7	1.41	2.08	2.48	**3.98**	0.4115	0.1218	0.2138	**0.0384**

ASVINGHTL	329154	*Ankrd44*	Db 9-mers	20.7	1.38	2.14	2.39	2.34	0.2167	0.0238	0.0807	0.1214
ATKYFTNRL	63959	*Slc29a1*	Kb 9-mers	4.5	1.16	1.47	1.69	2.74	0.4921	0.4318	0.0909	0.1408
ATLVFHNL	20848	*Stat3*	Kb 8-mers	1605.7	-1.07	1.19	1.18	1.48	0.8386	0.4682	0.4884	0.3984
ATQVYPKL	107976	*Bre*	Kb 8-mers	221.6	-1.12	1.13	1.32	2.29	0.7515	0.6575	0.4152	0.0517
AVLRYTKL	19718	*Rfc2*	Kb 8-mers	255.4	1.29	2.04	**3.36**	**5.31**	0.5114	0.3565	**0.0286**	**0.0201**
AVLSFSTRL	16430	*Stt3a*	Kb 9-mers	14.6	-1.32	1.12	1.88	1.52	0.5066	0.8057	0.2874	0.6301
AVVEFSRNV	51797	*Ctps*	Kb 9-mers	56.1	-1.05	1.80	1.48	2.08	0.8034	0.0424	0.0437	0.0055
AVYSYKRL	232339	*Ankrd26*	Kb 8-mers	9.6	1.04	1.25	1.95	**3.59**	0.8717	0.5435	0.0284	**0.0157**
EIVTFERL	14359	*Fxr1h*	Kb 8-mers	30.6	2.09	2.08	2.72	3.23	0.2728	0.1557	0.1499	0.2615
ESYSFEARM	67187	*Zmynd19*	Kb 9-mers	25.2	1.27	1.63	2.04	3.72	0.5697	0.2743	0.1029	0.1243
FAVVNHQGTL	216825	*Usp22*	Db 10-mers	317.1	-1.20	1.19	1.16	1.37	0.5470	0.4529	0.5885	0.3067
FAYRFSNL	110816	*Pwp2*	Kb 8-mers	188.6	1.28	1.34	1.51	1.77	0.6208	0.4391	0.2599	0.3095
FGPVNHEEL	12576	*Cdkn1b*	Db 9-mers	2064.7	-1.17	1.09	1.24	1.95	0.5371	0.6863	0.2499	0.0363
FLVQNIHTL	67832	*Bxdc2*	Db 9-mers	11.8	1.34	1.74	2.79	3.11	0.1296	0.0631	0.0547	0.1213
FQAINAGHI	66860	*Tanc1*	Db 9-mers	20.0	1.10	1.64	2.79	2.76	0.7442	0.1205	0.0528	0.0152
FSQENTEKI	76131	*Depdc1a*	Db 9-mers	12.4	-1.07	1.85	2.02	2.71	0.8922	0.2799	0.3108	0.0132
GAIVNGKVL	56784	*Garnl1*	Db 9-mers	25.5	1.14	1.27	1.17	1.50	0.6709	0.2149	0.6591	0.1198
GALENAKAEI	67463	*1200014M14Rik*	Db 10-mers	97.8	-1.04	1.66	1.47	2.09	0.9158	0.2149	0.3334	0.0089
GANVNHTTV	14155	*Fem1b*	Db 9-mers	38.4	-1.43	1.84	1.32	2.60	0.7008	0.4205	0.7683	0.4249
GAVKNLTYF	20443	*St3gal4*	Db 9-mers	38.1	1.03	2.06	1.96	5.66	0.9553	0.0948	0.1175	0.0569
GAVTNVKVI	15568	*Elavl1*	Db 9-mers	72.3	1.31	2.12	2.04	2.86	0.4475	0.0326	0.0307	0.0605
GAVTNVKVI	15569	*Elavl2*										
GAVTNVKVI	15571	*Elavl3*										
GGIQNVGHI	23918	*Impdh2*	Db 9-mers	129.2	-1.01	1.22	1.15	**1.66**	0.9798	0.5024	0.6675	0.1344
GGVVNMYHM	16913	*Psmb8*	Db 9-mers	1078.1	-1.01	1.38	2.61	**5.39**	0.9766	0.4627	0.1553	**0.0188**
GIMRFVNI	210544	*Wdr67*	Kb 8-mers	16.7	1.33	1.95	-1.29	3.30	0.6248	0.3595	0.6402	0.1916
GSLANHTSI	77573	*Vps33a*	Db 9-mers	34.9	-1.12	1.71	1.99	**3.90**	0.7933	0.2711	0.0954	**0.0083**
GSLKNVTTL	59079	*Erbb2ip*	Db 9-mers	85.8	1.28	2.03	2.16	**4.64**	0.5259	0.0692	0.0487	**0.0345**
GVLRFVNL	380959	*Alg10b*	Kb 8-mers	55.5	1.04	1.71	2.05	2.91	0.9485	0.4526	0.1389	0.4452
HAIENIDTF	12190	*Brca2*	Db 9-mers	155.4	-1.04	1.62	1.29	1.80	0.9131	0.0615	0.2626	0.1212
HQLQERHQL	20868	*Stk10*	Qa2 9-mers	18.1	-1.11	1.26	1.48	1.32	0.8395	0.6470	0.3733	0.6544
HTVQNADQV	21355	*Tap2*	Db 9-mers	7.2	1.55	2.21	3.44	2.55	0.1277	0.0737	0.1770	0.2952
IALRYVAL	70349	*Copb1*	Kb 8-mers	39.4	1.32	1.78	2.79	4.35	0.4979	0.1445	0.0016	0.1544
IFYYVQKL	207304	*Hectd1*	Kb 8-mers	23.3	1.35	2.05	2.34	2.39	0.5518	0.0375	0.0001	0.2970
IGLAYVNHL	218035	*Vps41*	Kb 9-mers	20.8	1.07	1.59	2.21	2.64	0.8365	0.2252	0.0750	0.3839

IGPRYSSV	320487	*Heatr5a*	Kb 8-mers	4.6	-1.17	1.30	1.27	3.11	0.6384	0.2408	0.2949	0.0552
IILKYIGM	54208	*Arl6ip1*	Kb 8-mers	36.7	1.53	1.77	5.82	19.53	0.6753	0.5212	0.0959	0.1771
IITGFRNV	12236	*Bub1b*	Kb 8-mers	240.6	-1.70	1.86	1.72	3.62	0.4874	0.2289	0.2435	0.0506
ILSVFPKV	241035	*Pkhd1*	Kb 8-mers	35.7	-1.10	1.89	**3.08**	7.09	0.8943	0.3478	**0.0484**	0.1779
IMVRNIDLV	20955	*Vamp7*	Db 9-mers	35.9	-1.08	1.38	1.67	2.24	0.8692	0.6288	0.4361	0.4366
INFDFPKL	13209	*Ddx6*	Kb 8-mers	70.9	-1.02	2.02	2.95	2.90	0.9835	0.5085	0.2661	0.4770
INIHYTQL	230895	*Vps13d*	Kb 8-mers	241.4	2.73	1.89	-11.02	6.75	0.5714	0.6579	0.3884	0.4402
INQIYEARV	216859	*Centb1*	Kb 9-mers	299.4	-1.05	1.49	1.77	**2.89**	0.8201	0.1192	0.0729	**0.0164**
IQLMNTAHL	74114	*Crot*	Db 9-mers	28.1	-1.01	1.60	2.51	4.73	0.9875	0.3579	0.3236	0.1242
IQVRNPVAL	72154	*Zfp157*	Db 9-mers	88.6	-1.39	1.54	1.90	3.06	0.4540	0.2024	0.0135	0.0273
ISGVNGTHI	20815	*Srpk1*	Db 9-mers	192.8	1.05	1.19	1.37	2.30	0.9129	0.5825	0.3513	0.0920
ISLEFRNL	14137	*Fdft1*	Kb 8-mers	264.2	-1.25	1.28	1.25	1.45	0.6568	0.5977	0.5257	0.7285
ISYLYNKL	22057	*Tob1*	Kb 8-mers	7.7	-1.23	-1.10	-1.12	1.37	0.5713	0.7469	0.7302	0.5154
ISYLYNKL	57259	*Tob2*										
KAFDYPSRL	212281	*A530054K11Rik*	Kb 9-mers	11.2	2.15	1.20	2.75	2.95	0.1337	0.7270	0.0774	0.1260
KAFDYPSRL	431706	*Zfp457*										
KAFDYPSRL	238690	*Zfp458*										
KAFDYPSRL	218314	*Zfp595*										
KAFTYINL	22042	*Tfrc*	Kb 8-mers	208.8	-1.18	1.25	1.26	1.42	0.6518	0.5012	0.4591	0.6968
KAIVNVIGM	81898	*Sf3b1*	Db 9-mers	28.5	1.21	1.33	1.82	2.13	0.6088	0.5355	0.1600	0.1205
KALINADEL	20740	*Spna2*	Db 9-mers	1472.3	-1.03	1.04	-1.03	1.44	0.9448	0.9051	0.9434	0.3486
KAPTFEVQM	73828	*Wdr21*	Kb 9-mers	8.4	1.67	2.93	**3.97**	**5.41**	0.1007	0.0237	**0.0180**	**0.0419**
KAVDFDGRL	207165	*Bptf*	Kb 9-mers	56.2	1.29	1.89	**3.20**	**4.70**	0.2519	0.2659	**0.0282**	**0.0043**
KAVENSSTAI	19167	*Psma3*	Db 10-mers	124.8	1.29	1.95	1.67	2.60	0.6273	0.0952	0.2042	0.0156
KGIGNKTEI	234135	*Whsc1l1*	Db 9-mers	328.9	-1.02	1.39	1.34	2.44	0.8791	0.4267	0.2252	0.2132
KIFEFKETL	227541	*Camk1d*	Kb 9-mers	28.4	-1.46	1.11	4.76	4.09	0.4616	0.8145	0.0862	0.3660
KIIEFANI	17427	*Mns1*	Kb 8-mers	25.9	-1.06	1.25	1.29	2.00	0.9050	0.1544	0.3287	0.3274
KILTFDQL	19899	*Rpl18*	Kb 8-mers	10.6	1.53	-1.12	2.16	3.69	0.2447	0.8241	0.1388	0.1826
KIQSFINRM	231874	*AU022870*	Kb 9-mers	137.6	1.33	1.61	2.64	3.40	0.6005	0.5399	0.0572	0.0878
KLHDVEHVL	56612	*Pfdn5*	Qa2 9-mers	91.9	-1.22	-1.04	1.16	1.30	0.4891	0.9356	0.7933	0.7148
KNFAFTLV	320571	*4930417M19Rik*	Kb 8-mers	7.7	1.17	1.50	2.67	2.40	0.7199	0.4144	0.0820	0.3118
KNFAFTLV	54670	*Atp8b1*										
KNFAFTLV	241633	*Atp8b4*										
KNLVNKEVM	14751	*Gpi1*	Db 9-mers	290.7	-1.27	-1.08	-1.24	1.45	0.1340	0.8272	0.0873	0.0771
KNVVYERV	69693	*Pof1b*	Kb 8-mers	268.8	-1.38	-1.06	1.41	1.67	0.0987	0.9128	0.0788	0.4451

KQLVNKEHL	51869	*Rif1*	Db 9-mers	190.6	-1.18	1.04	1.36	1.72	0.5598	0.9207	0.3274	0.0575
KQVNNLTNL	244879	*Npat*	Db 9-mers	146.6	1.23	1.51	2.53	2.44	0.7577	0.6732	0.0758	0.5018
KSFDYGNL	15273	*Hivep2*	Kb 8-mers	11.7	1.01	1.45	1.86	**2.94**	0.9666	0.2387	0.1313	**0.0286**
KSFLFSAL	170439	*Elovl6*	Kb 8-mers	80.5	-1.17	1.63	2.02	3.03	0.8314	0.4761	0.1842	0.4020
KSISNPPGSNL	232341	*Wnk1*	Db 11-mers	15.5	1.31	2.09	2.05	2.58	0.2727	0.0462	0.0119	0.0811
KSLQYLNL	16977	*Lrrc23*	Kb 8-mers	133.6	-1.00	1.50	1.77	2.37	0.9947	0.2551	0.0503	0.1021
KSYNFHTGL	74133	*1200011M11Rik*	Db 9-mers	19.2	-1.09	1.60	1.77	2.61	0.5373	0.3133	0.0196	0.2418
KSYSFIARM	27967	*Cherp*	Kb 9-mers	8.5	-1.04	1.02	2.09	2.56	0.8930	0.9680	0.0152	0.1167
KSYVNPTEL	228850	*B230339M05Rik*	Db 9-mers	14.7	-1.06	1.70	1.18	2.42	0.8308	0.2927	0.7386	0.0837
KTFLFSATM	67755	*Ddx47*	Kb 9-mers	21.2	1.31	1.29	1.82	1.51	0.5157	0.5849	0.1469	0.5225
KTFSFKSL	22640	*Zfp1*	Kb 8-mers	61.9	-1.23	-1.11	1.06	1.03	0.6625	0.7171	0.8726	0.9646
KTIYNVEHL	22644	*Rnf103*	Db 9-mers	60.5	-1.03	1.87	2.05	3.47	0.8943	0.2164	0.3447	0.0688
KTWRFSNM	218952	*Fermt2*	Kb 8-mers	42.7	1.05	2.34	2.53	4.56	0.8628	0.1443	0.0054	0.1542
KTWRFSNM	108101	*Fermt3*										
KVLVFSQM	15201	*Hells*	Kb 8-mers	184.9	1.11	1.28	1.37	1.54	0.7785	0.6531	0.3866	0.4222
KVVEFSEL	433864	*Gm1040*	Kb 8-mers	54.0	-1.06	1.42	1.48	2.24	0.8586	0.3265	0.1509	0.1262
KVVKFSYM	20747	*Spop*	Kb 8-mers	17.8	1.14	1.52	2.04	2.52	0.7072	0.2957	0.0509	0.0526
KVVKFSYM	76857	*Spopl*										
KVYNYNHL	19942	*Rpl27*	Kb 8-mers	420.4	1.00	1.65	1.67	**4.63**	0.9913	0.3498	0.1148	**0.0012**
KVYTFNSV	67972	*Atp2b1*	Kb 8-mers	15.0	-1.04	1.32	1.18	2.04	0.8992	0.3948	0.5863	0.1320
KVYTFNSV	11941	*Atp2b2*										
KVYTFNSV	320707	*Atp2b3*										
KVYTFNSV	381290	*Atp2b4*										
LGVTNFVHM	239554	*Foxred2*	Db 9-mers	12.1	-1.08	1.35	2.58	3.81	0.8460	0.5608	0.1184	0.2501
LNYKFPGL	106042	*Prickle1*	Kb 8-mers	284.4	1.00	1.35	1.95	3.25	0.9931	0.3353	0.0638	0.1516
LQYEFTHL	57778	*Fmnl1*	Kb 8-mers	474.0	1.01	1.38	1.70	2.47	0.9661	0.1982	0.0228	0.1429
LQYEFTKL	71409	*Fmnl2*										
LQYEFTKL	22379	*Fmnl3*										
LSPINHNTL	78757	*4921505C17Rik*	Db 9-mers	38.8	-1.08	1.53	1.67	2.06	0.7591	0.1519	0.1152	0.0159
LSVRNGATL	170826	*Ppargc1b*	Db 9-mers	58.3	-1.56	1.16	1.17	2.57	0.2906	0.6951	0.6398	0.1390
LSYSFAHL	195046	*Nlrp1a*	Kb 8-mers	5.2	1.08	1.03	1.53	5.10	0.8711	0.9418	0.3600	0.2849
LSYSFAHL	637515	*Nlrp1b*										
LSYSFAHL	627984	*Nlrp1c*										
LVYQFKEM	13709	*Elf1*	Kb 8-mers	311.5	-2.73	1.50	-2.10	2.79	0.3069	0.6539	0.4113	0.4814
LVYQFKEM	56501	*Elf4*										

MAHVNGVHL	382051	*4833426J09Rik*	Db 9-mers	198.3	-1.17	1.11	1.16	2.12	0.4866	0.7673	0.4276	0.0046
NAIKNHWNSTM	17863	*Myb*	Db 11-mers	163.7	1.04	1.66	2.24	**4.34**	0.9141	0.3200	0.0083	**0.0137**
NQFVNKFNVL	26893	*Cops6*	Db 10-mers	34.5	1.04	1.37	1.82	1.84	0.9684	0.7553	0.4664	0.5656
NQVKNIAEL	50505	*Ercc4*	Db 9-mers	12.8	1.03	1.36	1.32	2.17	0.8754	0.5578	0.4743	0.1740
NSFRYNGL	19943	*Rpl28*	Kb 8-mers	27.5	1.08	1.07	2.14	2.12	0.8582	0.8806	0.1278	0.1636
NSIRNLDTI	54138	*Atxn10*	Db 9-mers	123.1	-1.20	1.21	1.01	1.28	0.6510	0.5064	0.9664	0.3615
NTVTNKVTL	230700	*Foxj3*	Db 9-mers	30.1	1.39	1.99	2.70	**4.70**	0.2904	0.1499	0.0522	**0.0033**
NTYKYAKI	99382	*Abtb2*	Kb 8-mers	143.6	-1.31	1.23	1.69	**3.39**	0.2526	0.4505	0.0187	**0.0002**
QGPEYIERL	74016	*Phf19*	Kb 9-mers	17.7	1.74	2.15	2.81	3.32	0.2244	0.1115	0.3267	0.0513
QIIPFKTL	66409	*Rsl1d1*	Kb 8-mers	190.5	1.86	1.90	2.51	2.07	0.0707	0.2662	0.0959	0.2000
QLSDTLHSL	214901	*Chtf18*	Qa2 9-mers	24.7	1.00	1.56	1.64	**4.47**	0.9957	0.3943	0.0952	**0.0440**
QQLDVEHNL	66366	*Ergic3*	Qa2 9-mers	25.7	-1.62	1.92	2.25	**2.90**	0.5799	0.1000	0.0711	**0.0141**
QSVAFTKL	240641	*Mphosph1*	Kb 8-mers	25.1	1.01	1.85	1.97	**4.07**	0.9826	0.1691	0.2312	**0.0029**
QTVENVEHL	75425	*2610036D13Rik*	Db 9-mers	302.7	-1.08	1.74	2.08	2.99	0.8096	0.0568	0.1046	0.0089
QVVQFNRL	56692	*Map2k1ip1*	Kb 8-mers	155.6	4.10	5.11	1.38	6.04	0.1333	0.2068	0.3528	0.0892
QVYTFTERM	20448	*St6galnac4*	Kb 9-mers	18.0	1.02	1.24	1.55	2.97	0.9562	0.5646	0.2424	0.1170
RAFEFTYV	76355	*Tgds*	Kb 8-mers	27.4	2.05	2.82	1.92	3.06	0.2017	0.1921	0.4305	0.3119
RAIAFQHL	18432	*Mybbp1a*	Kb 8-mers	81.5	1.28	1.37	2.16	1.78	0.3769	0.2798	0.0289	0.2055
RALSNLESI	56085	*Ubqln1*	Db 9-mers	90.9	1.15	1.55	1.41	1.87	0.7478	0.0824	0.2147	0.0732
RALSNLESI	244178	*Ubqln3*										
RANQNFDEI	77480	*C330002I19Rik*	Db 9-mers	225.1	-1.35	-1.11	1.06	1.45	0.6839	0.8761	0.9220	0.5492
RAPAFHQL	268721	*2310021P13Rik*	Kb 8-mers	14.1	1.38	2.59	2.46	**4.61**	0.4377	0.1426	0.1017	**0.0177**
RAPQFINL	213109	*Phf3*	Kb 8-mers	59.4	-1.06	1.36	1.68	2.17	0.8216	0.2387	0.2033	0.1079
RAVDNQVYV	52633	*Nit2*	Db 9-mers	182.8	-1.02	1.66	1.38	**3.23**	0.9607	0.1842	0.3282	**0.0123**
RAVKNQIAL	56321	*Aatf*	Db 9-mers	80.9	-1.52	-1.04	-1.01	2.15	0.1045	0.8987	0.9452	0.0060
RAYLFNSV	106369	*Ypel1*	Kb 8-mers	18.1	1.19	1.84	2.67	3.88	0.2555	0.0960	0.0685	0.0548
RAYLFNSV	77864	*Ypel2*										
RAYLFNSV	66090	*Ypel3*										
RAYLFNSV	241525	*Ypel4*										
RNFIFSRL	15201	*Hells*	Kb 8-mers	640.3	-1.72	-1.30	2.16	1.69	0.3763	0.6340	0.1259	0.6335
RNLDYARL	20425	*Shmt1*	Kb 8-mers	241.9	-1.15	1.26	1.60	1.71	0.6296	0.5062	0.0719	0.4570
RQAENGMYI	69719	*Cad*	Db 9-mers	39.1	1.52	2.35	3.65	3.13	0.5407	0.3064	0.1054	0.2559
RQATNQIVM	19155	*Npepps*	Db 9-mers	80.0	1.06	1.32	1.97	1.60	0.8542	0.4249	0.0750	0.4031
RQILNADAM	66185	*1110037F02Rik*	Db 9-mers	97.7	1.01	1.30	1.43	1.70	0.9767	0.4200	0.2193	0.1481
RQLENGTTL	52635	*D12Ertd551e*	Db 9-mers	511.7	1.34	2.42	2.68	**3.68**	0.6597	0.0759	0.0626	**0.0042**

RSIQNAQFL	11432	*Acp2*	Db 9-mers	197.0	-1.26	1.35	1.38	1.67	0.1639	0.2515	0.0170	0.0278
RSISFSNM	18201	*Nsmaf*	Kb 8-mers	332.0	1.29	2.25	2.17	**3.47**	0.6761	0.0385	0.0711	**0.0036**
RSPENPPSKEL	66570	*Cenpm*	Db 11-mers	114.3	1.07	2.04	2.13	**3.90**	0.9157	0.3346	0.0844	**0.0129**
RTYSFLNL	230770	*Tmem39b*	Kb 8-mers	81.6	-1.32	1.53	1.84	1.87	0.5760	0.4818	0.2190	0.6002
RTYTYEKL	12387	*Ctnnb1*	Kb 8-mers	334.4	-1.41	-1.00	1.05	1.83	0.2602	0.9989	0.8318	0.0216
RVAEFTTNL	17886	*Myh9*	Kb -9mers	75.9	1.63	1.94	2.26	2.00	0.0670	0.2037	0.0157	0.2482
RVLIFSQM	12648	*Chd1*	Kb 8-mers	66.4	1.89	2.45	2.39	3.71	0.1940	0.1190	0.0364	0.0706
RVLIFSQM	244059	*Chd2*										
RVLIFSQM	216848	*Chd3*										
RVLIFSQM	107932	*Chd4*										
RVLIFSQM	269610	*Chd5*										
RVLIFSQM	320790	*Chd7*										
RVLIFSQM	93761	*Smarca1*										
RVLIFSQM	93762	*Smarca5*										
RVVANSEEI	11666	*Abcd1*	Db 9-mers	6.4	2.88	2.51	3.94	4.19	0.2040	0.1519	0.3591	0.2172
SAARFALL	107723	*Slc12a6*	Kb 8-mers	5.3	1.18	1.97	1.86	2.52	0.6970	0.1671	0.0884	0.0467
SAFSFRTL	107035	*Fbxo38*	Kb 8-mers	157.3	1.07	1.44	1.34	2.41	0.8517	0.1866	0.2559	0.2878
SALENAENHV	217216	*BC030867*	Db 10-mers	62.5	-1.06	1.46	1.53	2.48	0.8448	0.1849	0.3219	0.0247
SALRFLNL	71728	*Stk11ip*	Kb 8-mers	27.1	-1.01	1.48	2.00	2.57	0.9927	0.5427	0.1224	0.3554
SALRFQAM	74015	*Fcho1*	Kb 8-mers	56.4	1.40	1.86	**3.23**	**5.41**	0.2643	0.2905	**0.0229**	**0.0203**
SALVFTRL	56700	*0610031J06Rik*	Kb 8-mers	21.2	1.23	1.38	1.63	2.39	0.6292	0.2174	0.0854	0.3125
SAPENAVRM	14790	*Grcc10*	Db 9-mers	333.1	1.31	1.92	2.85	**4.42**	0.3092	0.0716	0.0790	**0.0059**
SAVIFRTL	326622	*Upf2*	Kb 8-mers	145.1	1.22	1.55	1.57	1.66	0.5829	0.1106	0.1465	0.2242
SAVKNLQQL	17168	*Mare*	Db 9-mers	1262.4	-1.21	1.15	1.08	1.52	0.6093	0.6098	0.7903	0.1586
SGLKYVNV	108143	*Taf9*	Kb 8-mers	685.2	-1.16	-1.01	-1.03	1.36	0.7775	0.9820	0.9398	0.4076
SGPTYIKL	57869	*Adck2*	Kb 8-mers	54.2	-1.01	1.40	1.47	2.26	0.9762	0.2050	0.2638	0.0382
SGVRFNNV	53321	*Cntnap1*	Kb 8-mers	5.5	1.45	2.51	2.37	**5.56**	0.2724	0.0414	0.0194	**0.0290**
SGVSNPHVI	208146	*Yeats2*	Db 9-mers	111.4	-1.03	1.43	1.45	2.49	0.9252	0.2152	0.2059	0.0125
SGYDFSRL	235623	*Scap*	Kb 8-mers	39.9	-1.17	1.14	1.19	2.08	0.6363	0.6391	0.4665	0.0965
SGYEFIHKL	70120	*Yars2*	Kb 9-mers	34.1	1.09	1.66	2.08	1.79	0.9220	0.5348	0.3211	0.6247
SGYIYHKL	19704	*Upf1*	Kb 8-mers	795.3	-1.05	1.18	1.48	1.48	0.8642	0.5879	0.1044	0.1061
SGYKFFSL	74781	*Wipi2*	Kb 8-mers	102.3	1.13	1.58	2.07	2.22	0.7883	0.3763	0.1097	0.4270
SGYKYVGM	68090	*Yif1a*	Kb 8-mers	80.7	1.39	2.78	3.41	**4.16**	0.4492	0.0445	0.0965	**0.0201**
SGYSFTHI	67019	*Actr6*	Kb 8-mers	170.6	1.00	1.59	1.36	2.17	0.9945	0.1071	0.1398	0.0670
SKYDFPKL	21822	*Tgtp*	Kb 8-mers	30.6	-1.18	1.29	2.59	1.77	0.7696	0.6586	0.0482	0.6127

SLGKNPTDAYL	67938	*Mylc2b*	Db 11-mers	31.1	1.21	1.40	2.65	**3.66**	0.6667	0.7010	0.1230	**0.0196**
SLITNKVVM	18950	*Pnp*	Kb 9-mers	51.6	1.25	1.78	3.34	**4.91**	0.7234	0.4052	0.1012	**0.0211**
SLLTNHVTL	15257	*Hipk1*	Db 9-mers	4.2	1.21	2.55	3.50	5.33	0.8191	0.2375	0.1001	0.1886
SLVKNVAHM	52521	*Zfp622*	Db 9-mers	120.9	-1.03	1.39	1.89	4.52	0.9591	0.6344	0.1175	0.0868
SMGKNPTDEYL	67268	*2900073G15Rik*	Db 11-mers	481.1	1.13	1.63	1.89	2.18	0.6725	0.2794	0.1670	0.0374
SMVQNRVFL	94176	*Dock2*	Db 9-mers	242.6	-1.07	1.09	1.31	1.41	0.8644	0.8447	0.4168	0.4173
SNFHFAVL	666173	*Vps13b*	Kb 8-mers	12.4	-1.09	1.46	1.46	1.86	0.8687	0.4322	0.3899	0.4910
SNIHYHTL	110749	*Chaf1b*	Kb 8-mers	40.6	1.36	2.06	2.04	2.62	0.4373	0.3211	0.0579	0.0283
SNLKYILV	29808	*Mga*	Kb 8-mers	45.0	-1.40	1.31	1.22	2.04	0.5274	0.6222	0.6401	0.3957
SNPEFRQL	338523	*Jhdm1d*	Kb 8-mers	6.7	1.12	1.58	2.08	**3.09**	0.6997	0.2976	0.0435	**0.0081**
SNVKYVML	19206	*Ptch1*	Kb 8-mers	9.8	1.05	1.26	1.50	3.55	0.8993	0.6915	0.5530	0.1132
SNYRVSLL	110078	*Pygb*	Kb 8-mers	11.8	1.14	1.65	1.76	2.52	0.5924	0.4275	0.2821	0.3518
SNYRVSLL	110095	*Pygl*										
SNYRVSLL	19309	*Pygm*										
SQAVNKQQI	12449	*Ccnf*	Db 9-mers	38.3	1.50	2.55	5.12	**6.05**	0.7097	0.2802	0.1295	**0.0247**
SQHVNLDQL	218978	*D14Ertd436e*	Db 9-mers	140.6	-1.04	1.50	2.18	**4.68**	0.8617	0.1550	0.0275	**0.0006**
SQLKNADVEL	11772	*Ap2a2*	Db 10-mers	18.5	1.10	1.58	3.05	**3.39**	0.7539	0.2023	0.0759	**0.0070**
SQLNELHHL	231123	*BC023882*	Qa2 9-mers	23.7	-1.34	1.09	1.12	1.22	0.3904	0.7710	0.6691	0.6199
SQLRNADVEL	11771	*Ap2a1*	Db 10-mers	7.5	-1.00	1.58	2.00	2.58	0.9991	0.1927	0.0172	0.0145
SQLRNEVAI	18760	*Prkcm*	Db 9-mers	57.7	1.10	1.66	1.92	2.73	0.7310	0.1985	0.0328	0.0068
SQLRNEVAI	75292	*Prkcn*										
SQLRNEVAI	101540	*Prkd2*										
SQVRNNVYM	A2A7B5	*Prdm2*	Db 9-mers	13.7	-1.11	-1.12	2.24	**3.20**	0.8374	0.8283	0.1389	**0.0449**
SQVRNNVYM	14411	*Slc6a12*										
SRVSFTSL	545216	*EG545216*	Kb 8-mers	26.9	1.67	2.18	**2.99**	**4.58**	0.1304	0.2481	**0.0226**	**0.0245**
SRVSFTSL	170772	*Glcci1*										
SSFQFHNRM	240066	*BC066107*	Kb 9-mers	25.1	1.13	1.21	1.99	**3.55**	0.5500	0.6940	0.0287	**0.0014**
SSIQNGKYTL	338523	*Jhdm1d*	Db 10-mers	303.9	-1.17	1.14	1.45	1.73	0.5943	0.6342	0.1183	0.0300
SSLINHKSV	22720	*Zfp62*	Db 9-mers	11.3	1.10	1.34	1.61	**3.13**	0.4941	0.2290	0.0080	**0.0007**
SSLQNHNHQL	109331	*Rnf20*	Db 10-mers	7.7	1.23	1.74	1.88	1.98	0.4924	0.4380	0.0373	0.0439
SSPSNKFFF	235626	*Setd2*	Db 9-mers	64.6	1.09	1.19	1.40	1.71	0.7901	0.6745	0.2693	0.2047
SSSLNQEKI	105203	*BC016423*	Db 9-mers	10.6	1.18	2.00	1.81	2.61	0.5601	0.0547	0.1056	0.0166
SSVQNYFHL	67095	*Trak1*	Db 9-mers	11.8	-1.10	1.57	1.87	2.20	0.7881	0.3203	0.0501	0.4828
SSVRFSYM	230082	*Nol6*	Kb 8-mers	9.3	1.19	1.49	1.49	2.76	0.6187	0.4256	0.3816	0.1172
SSVSNKTTL	227446	*2310035C23Rik*	Db 9-mers	49.8	-1.37	1.38	-1.05	1.63	0.4734	0.5745	0.8925	0.5486

SSVYFRSV	72368	*2310045N01Rik*	Kb 8-mers	8.8	1.54	2.32	2.14	**4.16**	0.4181	0.2975	0.1800	**0.0301**
SSYFFGKL	27405	*Abcg3*	Kb 8-mers	9.6	-1.13	1.30	**2.81**	5.71	0.7179	0.5120	**0.0064**	0.1395
SSYKFNHL	110606	*Fntb*	Kb 8-mers	32.2	-1.41	-1.12	1.19	1.59	0.2329	0.6922	0.4383	0.0722
SSYSFRHL	54397	*Ppt2*	Kb 8-mers	220.4	-1.44	1.27	-1.39	2.61	0.1917	0.5918	0.5000	0.1854
SSYTFPKM	23986	*Peci*	Kb 8-mers	74.1	-1.10	-1.04	1.19	1.47	0.8891	0.9574	0.7880	0.6033
STLIYRNM	75660	*Lin37*	Kb 8-mers	60.5	2.09	2.62	**3.37**	2.76	0.0742	0.0557	**0.0472**	0.0786
STLTYSRM	20393	*Sgk1*	Kb 8-mers	175.8	1.42	1.80	3.72	**9.21**	0.5974	0.3383	0.1758	**0.0321**
SVIKFENL	66980	*Zdhhc6*	Kb 8-mers	495.6	-1.07	1.06	1.10	1.10	0.8759	0.8621	0.7855	0.8180
SVMENSKVLGEAM	21894	*Tln1*	Db 13-mers	80.0	-1.09	-1.03	1.45	1.73	0.8555	0.9616	0.3807	0.3037
SVMENSKVLGESM	70549	*Tln2*	Db 13-mers	5.6	-1.76	1.07	1.18	1.17	0.1925	0.8754	0.6900	0.7036
SVVAFHNL	233115	*Dpy19l3*	Kb 8-mers	48.9	-1.02	1.31	1.63	2.42	0.9635	0.4235	0.1622	0.0501
SVYVYKVL	319178	*Hist1h2bb*	Kb 8-mers	309.1	1.04	1.84	1.87	1.99	0.9106	0.1176	0.0624	0.3626
SVYVYKVL	319181	*Hist1h2bg*										
SVYVYKVL	319182	*Hist1h2bh*										
SVYVYKVL	319183	*Hist1h2bj*										
SVYVYKVL	319184	*Hist1h2bk*										
SVYVYKVL	319186	*Hist1h2bm*										
SVYVYKVL	319188	*Hist1h2bp*										
SVYVYKVL	319189	*Hist2h2bb*										
SVYVYKVL	665596	*RP23-38E20.1*										
TAPQYYRL	67789	*Dalrd3*	Kb 8-mers	23.0	1.10	1.71	1.82	**2.82**	0.7153	0.0830	0.0343	**0.0188**
TGIKNGVHFL	50907	*Preb*	Db 10-mers	157.0	-1.08	1.38	2.29	2.27	0.8156	0.5469	0.0388	0.1757
TGPSNVDKL	12649	*Chek1*	Db 9-mers	102.7	-1.39	-1.20	1.23	1.23	0.5772	0.7503	0.5479	0.7227
TGVTNRDLI	71435	*Arhgap21*	Db 9-mers	50.9	1.26	-1.15	1.87	2.04	0.7046	0.7185	0.4233	0.2549
TIILFTKV	76582	*Ipo11*	Kb 8-mers	10.9	1.18	1.63	1.82	4.26	0.7763	0.3964	0.2753	0.4718
TNLRYLAL	11771	*Ap2a1*	Kb 8-mers	9.0	1.37	1.82	3.10	2.82	0.3156	0.3982	0.0866	0.1494
TNQDFIQRL	64652	*Nisch*	Kb 9-mers	335.0	1.54	2.05	3.37	**4.96**	0.5004	0.4254	0.1803	**0.0136**
TSIAFKNI	226778	*Mark1*	Kb 8-mers	19.6	1.20	2.11	1.88	2.63	0.5978	0.0129	0.0676	0.0188
TSIAFKNI	17169	*Mark3*										
TSIKNQTQL	78697	*Pus7*	Db 9-mers	206.7	1.08	1.44	1.69	1.65	0.8078	0.3835	0.1659	0.1284
TSVQFMKL	14104	*Fasn*	Kb 8-mers	10.6	-1.03	1.72	1.68	2.30	0.9394	0.2531	0.1903	0.2711
TSVRFTQL	66394	*Nosip*	Kb 8-mers	1435.0	-1.14	1.32	1.22	1.73	0.6878	0.2601	0.3980	0.0440
TSVVFNKL	66136	*Znrd1*	Kb 8-mers	5.9	1.11	1.41	1.67	**2.86**	0.7611	0.2998	0.1915	**0.0149**
TSYRFLAL	231123	*BC023882*	Kb 8-mers	270.7	1.10	1.50	1.65	2.27	0.7404	0.1888	0.1246	0.3012
TTYKYEMI	223691	*Eif3eip*	Kb 8-mers	26.8	2.12	1.97	3.00	3.26	0.1357	0.3330	0.0537	0.0877

TTYKYFAL	214290	*Zcchc6*	Kb 8-mers	55.5	1.34	1.03	1.92	1.46	0.3657	0.9417	0.0698	0.4963
VAAANREVL	574437	*Xlr3a/b*	Db 9-mers	907.2	1.05	1.78	1.88	**4.51**	0.8870	0.0876	0.0439	**0.0144**
VADKFSEL	73699	*Ppp2r1b*	Kb 8-mers	16.4	1.17	1.69	2.10	**3.37**	0.7307	0.3124	0.1775	**0.0290**
VAFAYKNV	67437	*Ssr3*	Kb 8-mers	130.7	-1.02	1.18	1.15	1.57	0.9525	0.5838	0.7000	0.1801
VAFDFTKV	233489	*Picalm*	Kb 8-mers	257.6	1.44	1.85	1.98	2.54	0.3410	0.0666	0.0324	0.2473
VAIGNPVHL	11906	*Zfhx3*	Db 9-mers	11.4	1.34	1.55	2.30	3.71	0.1898	0.2517	0.0182	0.1293
VALEFTHL	57913	*Lrdd*	Kb 8-mers	11.6	1.19	1.67	1.92	2.12	0.6665	0.2378	0.0483	0.4699
VAVKNSGGFL	52635	*D12Ertd551e*	Db 10-mers	21.5	-1.16	1.63	1.46	**4.45**	0.5712	0.1063	0.5969	**0.0034**
VAVTFSERL	26570	*Slc7a11*	Kb 9-mers	22.6	1.11	1.19	2.35	3.40	0.7134	0.2811	0.3609	0.3190
VAYGFRNI	67563	*Narfl*	Kb 8-mers	92.1	-1.11	1.05	-1.01	1.38	0.8080	0.8930	0.9710	0.4948
VEFFLDAL	320631	*Abca15*	Kb 8-mers	209.8	-1.01	1.75	1.34	2.21	0.9810	0.2248	0.3835	0.1858
VGIENIHVM	219135	*Mtmr6*	Db 9-mers	116.0	1.07	1.44	1.88	2.00	0.8217	0.2822	0.0798	0.1028
VGLINKDSV	192195	*Ash1l*	Db 9-mers	83.0	-1.06	1.54	1.56	**3.64**	0.8593	0.1421	0.2360	**0.0418**
VGPRYTNL	26413	*Mapk1*	Kb 8-mers	488.6	-1.11	1.05	-1.06	1.33	0.8176	0.8877	0.8806	0.3847
VGVTYRTL	212377	*F730047E07Rik*	Kb 8-mers	37.0	1.26	1.87	2.37	**4.78**	0.4871	0.0730	0.0521	**0.0189**
VGVYNGKTF	633683	*EG633683*	Db 9-mers	18.0	1.28	1.07	2.09	2.52	0.5222	0.8516	0.2556	0.0233
VGVYNGKTF	20054	*Rps15*										
VGYRFVTAI	64652	*Nisch*	Kb 9-mers	387.6	-1.07	1.73	1.54	2.15	0.8820	0.3090	0.2183	0.3932
VGYRYETL	109674	*Ampd2*	Kb 8-mers	249.2	-1.16	1.18	1.16	2.78	0.6810	0.6263	0.6442	0.0573
VINSFVHV	71745	*Cul2*	Kb 8-mers	25.9	1.01	1.23	1.57	1.83	0.9760	0.5337	0.1608	0.3565
VINVFHHL	56706	*Ccnl1*	Kb 8-mers	78.3	-1.14	1.26	2.21	2.74	0.6901	0.5402	0.0166	0.4355
VITEFARI	212528	*Trmt1*	Kb 8-mers	23.2	1.16	1.74	1.97	2.11	0.7275	0.1281	0.1156	0.2531
VITHFVHV	19395	*Rasgrp2*	Kb 8-mers	35.3	1.05	1.67	1.94	2.00	0.8257	0.1545	0.0515	0.1132
VITKFDHL	277939	*C2cd3*	Kb 8-mers	15.2	1.11	1.58	2.32	3.21	0.8300	0.5037	0.1576	0.1189
VITKFINV	240168	*Rasgrp3*	Kb 8-mers	15.6	1.43	1.98	**3.35**	4.78	0.5932	0.1360	**0.0481**	0.2140
VIVRFLTV	267019	*Rps15a*	Kb 8-mers	264.9	-1.24	1.38	1.18	1.49	0.6211	0.4127	0.7105	0.6835
VIVRFLTVM	267019	*Rps15a*	Kb 9-mers	23.9	1.40	1.37	1.59	1.65	0.5527	0.5629	0.2214	0.6660
VMLENYSHL	408067	*9630025I21Rik*	Db 9-mers	17.6	1.21	1.23	1.55	2.27	0.5412	0.5546	0.2837	0.2163
VMLENYSHL	67911	*Zfp169*										
VMLENYSHL	22693	*Zfp30*										
VMLENYSHL	245368	*Zfp300*										
VMLENYSHL	328274	*Zfp459*										
VMLENYSHL	78251	*Zfp712*										
VMLENYSHL	235907	*Zfp71-rs1*										
VMLENYSHL	212276	*Zfp748*										

VMLENYSHL	233056	*Zfp790*									
VMLENYSHL	238693	*Zfp817*									
VMLENYSHL	170763	*Zfp87*									
VMLENYSHL	22754	*Zfp92*									
VNFVHTNL	67891	*Rpl4*	Kb 8-mers	1167.5	-1.10	-1.27	-1.51	1.48	0.7969	0.6067	0.1625	0.1470
VNIEFKDL	11307	*Abcg1*	Kb 8-mers	306.4	1.16	1.70	2.44	2.25	0.6968	0.2153	0.0789	0.4521
VNLIYHQRI	212569	*Zfp273*	Kb 9-mers	5.3	1.48	2.78	**4.73**	4.68	0.4866	0.2702	**0.0131**	0.2525
VNLIYHQRI	22746	*Zfp85-rs1*										
VNQKFNNL	218977	*Dlg7*	Kb 8-mers	648.4	-1.27	1.46	1.51	2.69	0.3606	0.2345	0.1305	0.0007
VNRVFDKL	16912	*Psmb9*	Kb 8-mers	531.1	-1.11	1.35	2.73	**4.67**	0.5409	0.3286	0.0044	**0.0110**
VNVDYSKL	17992	*Ndufa4*	Kb 8-mers	752.5	-1.02	1.21	1.46	2.54	0.9580	0.3563	0.1805	0.0344
VNVRFTGV	75007	*4930504E06Rik*	Kb 8-mers	16.8	-1.11	1.43	1.82	1.87	0.8160	0.3196	0.3262	0.2349
VNVRFTGV	235461	*B230380D07Rik*										
VNYRHLAL	20020	*Polr2a*	Kb 8-mers	133.4	-1.38	1.15	1.79	2.39	0.2420	0.7217	0.0326	0.3175
VNYYFERNM	75725	*Phf14*	Kb 9-mers	12.9	1.30	1.77	2.53	4.03	0.5166	0.4967	0.1368	0.0547
VQYKFSHL	330129	*EG330129*	Kb 8-mers	743.3	-1.32	1.03	1.16	1.75	0.2411	0.9152	0.4324	0.1300
VQYKFSHL	14534	*Gcn5l2*										
VQYKFSHL	18519	*Pcaf*										
VSFTYRYL	80743	*Vps16*	Kb 8-mers	878.0	-1.05	1.27	1.21	1.59	0.9101	0.4363	0.3976	0.5249
VSILNRQVL	101706	*Numa1*	Db 9-mers	284.5	-2.22	1.73	2.32	2.29	0.0387	0.2559	0.0170	0.1039
VSLKYAHM	11566	*Adss*	Kb 8-mers	3.8	1.63	1.93	**3.26**	2.83	0.1793	0.0958	**0.0417**	0.0531
VSPQNVHHSYL	100040505	*LOC100040505*	Db 11-mers	96.0	-1.10	1.12	1.23	1.25	0.7642	0.7239	0.4419	0.6494
VSQYYPKL	15441	*Hp1bp3*	Kb 8-mers	24.8	1.09	1.36	1.96	**3.65**	0.8183	0.4273	0.1688	**0.0347**
VSTKFEHL	320714	*D030016E14Rik*	Kb 8-mers	57.1	-1.00	1.81	1.74	2.61	0.9918	0.1831	0.0337	0.0028
VSVEYTEKM	14791	*Emg1*	Kb 9-mers	336.1	1.25	1.74	2.01	2.09	0.4914	0.1516	0.1645	0.0944
VSYKYSKV	16004	*Igf2r*	Kb 8-mers	6.9	-1.48	-1.22	1.12	**3.18**	0.3191	0.5580	0.6924	**0.0018**
VSYLFSHV	328099	*AU021838*	Kb 8-mers	70.0	1.00	1.51	1.64	2.28	0.9857	0.1501	0.1240	0.2401
VSYLFSHV	19139	*Prps1*										
VSYLFSHV	75456	*Prps1l1*										
VSYLFSHV	110639	*Prps2*										
VSYNHTNI	382252	*A830080D01Rik*	Kb 8-mers	23.6	1.43	1.86	2.17	**3.63**	0.3444	0.0369	0.0492	**0.0007**
VTIHYNKL	18693	*Pick1*	Kb 8-mers	380.4	-1.28	-1.00	1.15	1.38	0.3388	0.9888	0.4835	0.1387
VVFQFQHI	12767	*Cxcr4*	Kb 8-mers	154.4	1.01	1.56	1.63	2.67	0.9782	0.1757	0.0716	0.3043
VVYIYHSL	67014	*Mina*	Kb 8-mers	34.4	-1.08	3.26	1.52	3.96	0.8895	0.3668	0.3589	0.1395
WAVSNREML	68477	*Rmnd5a*	Db 9-mers	45.5	1.50	2.17	**4.24**	**5.55**	0.1381	0.2364	**0.0046**	**0.0056**

YAIKNIHGV	13430	*Dnm2*	Db 9-mers	22.2	1.03	1.62	1.72	2.27	0.8732	0.3768	0.0151	0.0433
YAMENTRQTI	20185	*Ncor1*	Db 10-mers	5.6	1.15	1.33	2.19	2.41	0.6147	0.3761	0.2307	0.0475
YAMIYRNL	17246	*Mdm2*	Kb 8-mers	8.5	1.23	1.98	1.94	10.37	0.6040	0.1530	0.1146	0.2310
YGIQNHHEV	320817	*Atad2b*	Db 9-mers	28.6	-1.33	-1.08	1.25	1.85	0.3226	0.7903	0.5143	0.0579
YKNVNQEVV	53415	*Htatip2*	Db 9-mers	76.1	-1.23	1.19	1.27	1.76	0.5719	0.6952	0.4994	0.0426
YMGTNIHSL	16329	*Inpp1*	Db 9-mers	45.9	-1.05	1.21	1.49	1.82	0.8851	0.6533	0.3090	0.1481
YQKENKDVI	17420	*Mnat1*	Db 9-mers	245.2	-1.37	1.65	2.01	**3.56**	0.6944	0.4308	0.1314	**0.0113**
YSGLNQRAV	12501	*Cd3e*	Db 9-mers	26.4	1.08	1.44	1.78	2.16	0.7279	0.2700	0.1112	0.0165
YSVANHNSFL	74734	*Rhoh*	Db 10-mers	619.7	-1.16	1.35	1.35	1.62	0.5242	0.3083	0.0664	0.0769
YTVANKEYV	67458	*Ergic1*	Db 9-mers	52.9	-1.17	1.28	1.17	1.97	0.5902	0.3847	0.5104	0.0360
YVLHNSNTM	432448	*EG432448*	Db 9-mers	8.4	1.53	1.83	2.51	**3.46**	0.4058	0.3802	0.2187	**0.0167**
YVLHNSNTM	21762	*Psmd2*										

overexpressed MHC I peptides only detected 48 hours following rapamycin exposure

Peptide	Gene ID	Gene Symbol	Restriction size	WT/B2M*	FC T6/T0	FC T12/T0	FC T24/T0	FC T48/T0	P T6/T0	P T12/T0	P T24/T0	P T48/T0
KALSYASL	78757	*4921505C17Rik*	Kb 8-mers	9.7	#	#	#	**9.73**	#	#	#	**0.0026**
QQIAFKNL	319491	*1110029I05Rik*	Kb 8-mers	7.9	#	#	#	**7.92**	#	#	#	**0.0043**
QQIAFKNL	21781	*Tfdp1*										
QQIAFKNL	211586	*Tfdp2*										
VNSNFYLRM	68845	*Pih1d1*	Kb 9-mers	9.0	#	#	#	**8.96**	#	#	#	**0.0102**
VNTHFSHL	73668	*Ttc21b*	Kb 8-mers	15.3	#	#	#	**15.25**	#	#	#	**0.0003**
VSVTNEHLM	22035	*Tnfsf10*	Db 9-mers	9.0	#	#	#	**8.97**	#	#	#	**0.0072**
VSYKNPSLM	114741	*Supt16h*	Db 9-mers	8.9	#	#	#	**8.91**	#	#	#	**0.0114**

* Fold change ≥ 3 to < 5 corespond to low abundance MHC I peptides only detected in the WT condition

significantly differentially expressed H2Db peptides (FC ≤ - 1.9 and ≥ 3.1, P ≤ 0.05)

significantly differentially expressed H2Kb peptides (FC ≤ - 2.2 and ≥ 2.8, P ≤ 0.05)

significantly differentialy expressed Qa2 & Qa1 peptides (FC ≤ -2.7 and ≥ 2.3, P ≤ 0.05)

Table AIII.III: Functional classification of differentially expressed MHC I-associated peptides following mTOR inhibition on EL4 cells.

Functional classification	Gene symbol	Gene ID	Function	Sequence	Rapa/ CTRL	P value
Cell cycle	*Ccnf*	12449	Cyclin F, G2/M transition	SQAVNKQQI	6.0	0.02
	Cenpm	66570	Mitotic cell cycle progression	RSPENPPSKEL	3.9	0.01
	Lin37	75660	Member of the DREAM complex	STLIYRNM	3.4	0.05
	Mnat1	17420	Component of CAK, G1/S transition	YQKENKDVI	3.6	0.01
	Mphosph1	240641	M phase phosphoprotein 1	QSVAFTKL	4.1	0.003
	Tfdp1 / Tfdp2	21781 / 211586	E2F1/DP1, G1/S transition	QQIAFKNL*	7.9	0.004
	Pin1	23988	Regulation of cell proliferation	RVYYFNHI	3.0	0.02
Chromatin related	*Ash1l*	192195	Histone methyltransferase	VGLINKDSV	3.6	0.04
	Bptf	207165	Chromatin-mediated regulation of transcription	KAVDFDGRL	4.7	0.004
	Chtf18	214901	PCNA loader, DNA replication	QLSDTLHSL	4.5	0.04
	D14Ertd436e	218978	Chromosome segragation ATPases	SQHVNLDQL	4.7	0.001
	Hp1bp3	15441	Linker histone 1 and histone 5 domains	VSQYYPKL	3.7	0.03
	Rfc2	19718	PCNA loader, DNA replication	AVLRYTKL	5.3	0.02
	Supt16h	114741	Chromatin-mediated regulation of transcription	VSYKNPSLM*	8.9	0.01
Transcription	*BC066107*	240066	KRAB domain, Zn finger	SSFQFHNRM	3.5	0.001
	Foxj3	230700	Transcription factor	NTVTNKVTL	4.7	0.003
	Hivep2	15273	Transcription factor, role in lymphocyte development	KSFDYGNL	2.9	0.03
	Jhdm1d	338523	Histone demethylase	SNPEFRQL	3.1	0.008
	Myb	17863	Regulates proliferation of hematopoietic cells	NAIKNHWNSTM	4.3	0.01
	Taf6	21343	Subunit of TIIFD	AQQVNRTTL	4.1	0.003
	Zfp273/Zfp85-rs1	212569 / 22746	KRAB box, Zn-finger	VNLIYHQRI	4.7	0.01
	Zfp62	22720	Zn-finger	SSLINHKSV	3.1	0.001
	Znrd1	66136	Component of the RNA polymerase I (Pol I)	TSVVFNKL	2.9	0.01
Signal transduction	*4921505C17Rik*	78757	Component of mTORC2	KALSYASL*	9.7	0.003
	D12Ertd551e	52635	Protein Kinase C conserved region (CalB)	VAVKNSGGFL	4.4	0.003
	D12Ertd551e	52635	Protein Kinase C conserved region (CalB)	RQLENGTTL	3.7	0.004
	Erbb2ip	59079	Binding partner of Erbb2	GSLKNVTTL	4.6	0.03
	Nisch	64652	Rac protein signal transduction	TNQDFIQRL	5.0	0.01

	Ppp2r1b	73699	Regulates numerous signaling pathways	VADKFSEL	3.4	0.03
	Rasgrp3	240168	Ras protein signal transduction	VITKFINV	3.3	0.05
	Wdr21	73828	WD40 domain	KAPTFEVQM	5.4	0.04
Cell differentiation	*Dhcr7*	13360	Role in blood vessel development	AGVVNKYEV	3.8	0.009
	Glcci1	545216/170772	Role in hematopoiesis	SRVSFTSL	4.6	0.02
	Igf2r	16004	Role in embryonic development and growth	VSYKYSKV	3.2	0.002
	Pkhd1	241035	Cell-cell adhesion, kidney development	ILSVFPKV	3.1	0.05
	Xlr3a/b	574437	Role in lymphocyte development	VAAANREVL	4.5	0.01
Apoptosis	*Nsmaf*	18201	TNF receptor adaptor protein	RSISFSNM	3.5	0.004
	Sgk1	20393	Response to DNA damage stimulus	STLTYSRM	9.2	0.03
	Tnfsf10	22035	TNF-related apoptosis inducing ligand (TRAIL)	VSVTNEHLM*	9.0	0.007
Immune response	*H2-D1/H2-Q10/H2-Q5*	14964/15007/15016	Antigen processing and presentation	AMAPRTLLL	2.9	0.04
	Psmb8	16913	Antigen processing and presentation	GGVVNMYHM	5.4	0.02
	Psmb9	16912	Antigen processing and presentation	VNRVFDKL	4.7	0.01
Metabolism	*Adss*	11566	Purine nucleotide biosynthetic process	VSLKYAHM	3.3	0.04
	Ankrd26	232339	Negative regulation of multicellular organism growth	AVYSYKRL	3.6	0.02
	Pnp	18950	Purine nucleoside metabolic process	SLITNKVVM	4.9	0.02
Translation	*Pabpc1*	18458	mRNA processing	AGVRNPQQHL	3.4	0.001
	Rpl27	19942	Structural constituent of ribosome	KVYNYNHL	4.6	0.001
Transport	*Abcg3*	27405	ATP-binding cassette transporters	SSYFFGKL	2.8	0.006
	Ap2a2	11772	Vesicle-mediated transport	SQLKNADVEL	3.4	0.007
	Sec24c	218811	Vesicle-mediated transport	KVLHFYNV	3.3	0.002
	Vps33a	77573	Vesicle-mediated transport	GSLANHTSI	3.9	0.008
	Yif1a	68090	Vesicle-mediated transport	SGYKYVGM	4.2	0.02
Cytoskeleton	*Mylc2b*	67938	Cytoskeleton organization and biogenesis	SLGKNPTDAYL	3.7	0.02
	Tmod1	21916	Cytoskeleton organization	SSIVNKEGL	-3.7	0.006
Miscellaneous	*Abtb2*	99382	Protein binding	NTYKYAKI	3.4	0.0002
	Centb1	216859	GTPase-activating protein	INQIYEARV	2.9	0.02
	Cntnap1	53321	Intracellular processing and transport of contactin	SGVRFNNV	5.6	0.03
	Psmd2	21762	Regulation of protein catabolic process	YVLHNSNTM	3.5	0.02
	Nae1	234664	Ubiquitin-like protein NEDD8	AAVGNHVAKL	4.6	0.0003
	Nit2	52633	Role in cell growth inhibition	RAVDNQVYV	3.2	0.01
	Ormdl2	66844	Protein folding	AMYIFLHTV	3.8	0.006
	Rmnd5a	68477	Hypothetic role in cell division	WAVSNREML	5.5	0.006

Unknown	*2310021P13Rik*	268721	Unknown	RAPAFHQL	4.6	0.02
	2310045N01Rik	72368	Unknown	SSVYFRSV	4.2	0.03
	2810405J04Rik	72722	Unknown	ASPEFTKL	4.0	0.04
	A830080D01Rik	382252	Unknown	VSYNHTNI	3.6	0.001
	BC037112	231128	Unknown	AAPRNSPTGL	3.4	0.007
	Dalrd3	67789	Unknown	TAPQYYRL	2.8	0.02
	Ergic3	66366	Unknown	QQLDVEHNL	2.9	0.01
	F730047E07Rik	212377	Unknown	VGVTYRTL	4.8	0.02
	Fcho1	74015	Unknown	SALRFQAM	5.4	0.02
	Grcc10	14790	Unknown	SAPENAVRM	4.4	0.006
	Pih1d1	68845	Unknown	VNSNFYLRM*	9.0	0.01
	Ttc21b	73668	Unknown	VNTHFSHL*	15.3	0.0003

*Only detected following rapamycin exposure