

Université de Montréal

**Étude de la voie de signalisation de l'insuline chez la
drosophile par une approche phosphoprotéomique**

par

Gaëlle Bridon

Département de Biochimie

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de Doctorat
en Biochimie

Avril, 2012

© Gaëlle Bridon, 2012

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

Étude de la voie de signalisation de l'insuline
chez la drosophile par une approche phosphoprotéomique

présentée par:

Gaëlle Bridon

a été évaluée par un jury composé des personnes suivantes:

Dr. Benoit Coulombe, président-rapporteur

Dr. Pierre Thibault, directeur de recherche

Dr. Pierre Chaurand, membre du jury

Dr. Stéphane Angers, examinateur externe

Dr. Jacques Archambault, représentant du doyen de la FES

Résumé

La phosphorylation est une modification post-traductionnelle modulant l'activité, la conformation ou la localisation d'une protéine et régulant divers processus. Les kinases et phosphatases sont responsables de la dynamique de phosphorylation et agissent de manière coordonnée. L'activation anormale ou la dérégulation de kinases peuvent conduire au développement de cancers ou de désordres métaboliques. Les récepteurs tyrosine kinase (RTKs) sont souvent impliqués dans des maladies et la compréhension des mécanismes régissant leur régulation permet de déterminer les effets anticipés sur leurs substrats.

Dans ce contexte, le but de cette thèse est d'identifier les événements de phosphorylation intervenant dans la voie de l'insuline chez la drosophile impliquant un RTK : le récepteur de l'insuline (InR). La cascade de phosphorylation déclenchée suite à l'activation du récepteur est conservée chez le mammifère. Afin d'étudier le phosphoprotéome de cellules S2 de drosophile, nous avons utilisé une étape d'enrichissement de phosphopeptides sur dioxyde de titane suivie de leur séparation par chromatographie liquide (LC) et mobilité ionique (FAIMS). Les phosphopeptides sont analysés par spectrométrie de masse en tandem à haute résolution. Nous avons d'abord démontré les bénéfices de l'utilisation du FAIMS comparativement à une étude conventionnelle en rapportant une augmentation de 50 % dans le nombre de phosphopeptides identifiés avec FAIMS. Cette technique permet de séparer des phosphoisomères difficilement distinguables par LC et l'acquisition de spectres MS/MS distincts où la localisation précise du phosphate est déterminée. Nous avons appliqué cette approche pour l'étude des phosphoprotéomes de cellules S2 contrôles ou traitées à l'insuline et avons identifié 32 phosphopeptides (sur 2 660 quantifiés) pour lesquels la phosphorylation est modulée. Étonnamment, 50 % des cibles régulées possèdent un site consensus pour la kinase CK2. Une stratégie d'inhibition par RNAi a été implémentée afin d'investiguer le rôle de CK2 dans la voie de l'insuline. Nous avons identifié 6 phosphoprotéines (CG30085, su(var)205, scny, protein CDV3 homolog, D1 et mu2) positivement régulées suite à l'insuline et négativement modulées après le traitement par

RNAi CK2. Par essai kinase *in vitro*, nous avons identifié 29 cibles directes de CK2 dont 15 corrélaient avec les résultats obtenus par RNAi. Nous avons démontré que la phosphorylation de su(var)205 (S₁₅) était modulée par l'insuline en plus d'être une cible directe de CK2 suite à l'expérience RNAi et à l'essai kinase.

L'analyse des données phosphoprotéomiques a mis en évidence des phosphopeptides isomériques dont certains étaient séparables par FAIMS. Nous avons déterminé leur fréquence lors d'études à grande échelle grâce à deux algorithmes. Le script basé sur les différences de temps de rétention entre isomères a identifié 64 phosphoisomères séparés par LC chez la souris et le rat (moins de 1 % des peptides identifiés). Chez la drosophile, 117 ont été répertoriés en combinaison avec une approche ciblée impliquant des listes d'inclusion. Le second algorithme basé sur la présence d'ions caractéristiques suite à la fragmentation de formes qui co-éluent a rapporté 23 paires isomériques. L'importance de pouvoir distinguer des phosphoisomères est capitale dans le but d'associer une fonction biologique à un site de phosphorylation précis qui doit être identifié avec confiance.

Mots-clés : Spectrométrie de masse • FAIMS • Insuline • *Drosophila melanogaster* • Cellules S2 • Phosphoprotéomique • Phosphorylation • Protéomique quantitative • Signalisation cellulaire • Caséine kinase 2.

Abstract

Phosphorylation is a reversible post-translational modification that modulates protein activity, and can impart conformational changes and affect translocation of their protein substrates. Kinases and phosphatases are responsible for the dynamic of changes in protein phosphorylation and act in a coordinated manner. Abnormal activation or misregulation of kinase activity can lead to the development of cancers and metabolic disorders. Tyrosine kinase receptor (RTK) associated signaling pathways are often implicated in numerous diseases and the further understanding of mechanisms affecting their regulation is necessary to determine their activity and effects anticipated on their substrates.

In this context, the primary objective of this thesis is to study the phosphorylation events arising from the activation of the insulin receptor (InR) following stimulation of drosophila S2 cells with insulin. The phosphorylation cascade triggered after InR activation is conserved in mammals. In order to study the phosphoproteome of drosophila S2 cells, we enriched phosphopeptides on titanium dioxide (TiO₂) stationary phase prior to their separation by liquid chromatography (LC) and ion mobility (FAIMS) mass spectrometry (MS). Phosphopeptides were then analysed by tandem MS at high resolution. We first compared the benefits of FAIMS to conventional LC-MS, and observed a 50% increase in the number of identified phosphopeptides when using ion mobility. FAIMS enables the separation of phosphoisomers that are typically unresolved by LC, enabling high confidence assignment of modification sites via distinct MS/MS spectra. This approach was used to profile phosphorylation changes taking place between control and insulin-treated drosophila cells and enabled the identification of 32 phosphopeptides (out of 2 660 quantified) showing differential regulation. Interestingly, 50% of the regulated targets have a CK2 consensus site. These preliminary experiments were followed-up by RNAi mediated inhibition of CK2 and revealed that 6 phosphoproteins (CG30085, su(var)205, scny, protein CDV3 homolog, D1 and mu2) were positively modulated after insulin stimulation and negatively regulated after CK2 RNAi treatment. Using *in vitro* kinase assay, we identified 29 direct CK2 targets, of which 15 were correlated with results from the CK2 RNAi

experiment. We demonstrated specifically that the su(var)205 (S₁₅) is regulated by insulin and is a direct CK2 target based on RNAi and kinase assays.

Our phosphoproteomics data also highlighted the presence of isomeric phosphopeptides, several of which could be distinguished using FAIMS. We developed two algorithms to determine the occurrence of phosphoisomers in large scale studies. The first algorithm based on differences in retention times between isomers identified 64 candidates in mouse and rat phosphoproteome datasets corresponding to less than 1% of all identified phosphopeptides. We also identified 117 isomer candidates in drosophila using a targeted LC-MS/MS approach with inclusion lists. The second algorithm is based on the presence of characteristic fragment ions present in MS/MS spectra of co-eluting or partially resolved species and allowed the identification of 23 isomeric pairs. The ability to distinguish phosphoisomers in large-scale phosphoproteome datasets is of significance to correlate phosphorylation events taking place on specific residues with biological activities.

Keywords : Mass spectrometry • FAIMS • Insulin • *Drosophila melanogaster* • S2 cells • Phosphoproteomics • Phosphorylation • Quantitative proteomics • Signaling pathway • Casein kinase 2.

Table des matières

RÉSUMÉ	I
ABSTRACT	III
TABLE DES MATIÈRES	V
LISTE DES TABLEAUX	XII
LISTE DES FIGURES	XIII
LISTE DES ABRÉVIATIONS	XVII
REMERCIEMENTS	XXV
CHAPITRE 1 : INTRODUCTION	1
1.1. Organisation des chapitres	2
1.2. Les protéines	5
1.3. La phosphorylation	6
1.3.1. Mécanisme général.....	6
1.3.2. Kinases et phosphatases	6
1.3.3. La phosphoprotéomique	12
1.3.4. L'intérêt clinique de l'étude de la phosphorylation.....	13
1.4. La drosophile comme modèle d'étude	15
1.4.1. Les cellules S2 de drosophile	16
1.4.2. La stratégie RNAi dans les cellules S2	16
1.5. La voie de signalisation de l'insuline	19
1.5.1. Le rôle de l'insuline dans le métabolisme	19

1.5.2.	L'activation de InR et la cascade de phosphorylation en aval	21
1.5.2.1.	La voie Ras-MAPK.....	23
1.5.2.2.	La voie de PI3K et AKT	24
1.5.2.3.	La régulation du taux de glucose via GLUT4	26
1.5.3.	Les études phosphoprotéomiques de la voie de l'insuline	27
1.6.	Protéine kinase CK2.....	28
1.7.	La préparation d'échantillon en phosphoprotéomique.....	30
1.7.1.	La lyse cellulaire.....	32
1.7.2.	Digestion enzymatique	32
1.7.3.	Détection de phosphoprotéines par électrophorèse sur gel.....	33
1.7.3.1.	Gel SDS-PAGE en une dimension.....	33
1.7.3.2.	Immunobuvardage de type Western et approches utilisant des anticorps	34
1.7.3.3.	Coloration spécifiques aux phosphoprotéines	35
1.7.3.4.	Autoradiographie	36
1.7.4.	Les méthodes d'enrichissement en phosphoprotéines et phosphopeptides.....	36
1.7.4.1.	Immunoprécipitation à l'aide d'anticorps	38
1.7.4.2.	Chromatographie d'affinité au métal immobilisé (IMAC).....	38
1.7.4.3.	Le dioxyde de titane	39
1.7.4.4.	L'approche SIMAC (élution séquentielle après IMAC)	40
1.7.4.5.	Capture par affinité aux ions métalliques basés sur un polymère (PolyMAC)	41
1.7.4.6.	Autres chromatographies d'affinité pour l'enrichissement en phosphopeptides	42
1.7.4.7.	Précipitation au phosphate de calcium	42
1.7.4.8.	Dérivatisation chimique	43
1.8.	Séparation par chromatographie liquide	45
1.8.1.	Séparation selon l'hydrophobicité	45
1.8.2.	Séparation par échange d'ions	45
1.9.	La spectrométrie de masse comme outil d'étude de la phosphorylation	47
1.9.1.	Composantes d'un spectromètre de masse	48
1.9.1.1.	Sources d'ionisation.....	48
1.9.1.2.	Analyseurs de masse	50
1.9.1.3.	Détecteurs.....	56

1.9.2.	Spectrométrie de mobilité ionique modulée par un champ électrique en forme d'onde asymétrique à champ élevé (FAIMS)	57
1.9.3.	Spectrométrie de masse en tandem : méthodes de fragmentation	61
1.9.3.1.	CID (dissociation induite par collision).....	62
1.9.3.2.	ETD (dissociation par transfert d'électrons).....	66
1.9.3.3.	HCD (dissociation induite par collision de haute énergie)	68
1.9.4.	Méthodes de quantification	69
1.9.4.1.	Quantification sans marquage.....	70
1.9.4.2.	Quantification par marquage métabolique.....	73
1.9.4.3.	Quantification par marquage chimique.....	75
1.10.	Bioinformatique appliquée aux études protéomiques.....	77
1.10.1.	Traitement des données brutes et interrogation des bases de données	77
1.10.2.	Outils bioinformatiques dédiés à la phosphoprotéomique.....	79
1.10.3.	Logiciel Proteoprofile pour la quantification sans marquage	82
1.11.	Objectifs de recherche	85
CHAPITRE 2 : IMPROVEMENT OF PHOSPHOPROTEOME ANALYSES USING FAIMS AND DECISION TREE FRAGMENTATION; APPLICATION TO THE INSULIN SIGNALLING PATHWAY IN DROSOPHILA MELANOGASTER S2 CELLS.....		
87		
2.1.	Contribution des auteurs.....	88
2.2.	Abstract	89
2.3.	Introduction.....	90
2.4.	Materials and methods	93
2.4.1.	Materials	93
2.4.2.	Cell culture and protein extraction	94
2.4.3.	Immunoblotting analysis	94
2.4.4.	Trypsin digestion.....	94
2.4.5.	Phosphopeptide isolation	95
2.4.6.	Liquid chromatography	96

2.4.7.	FAIMS interface and mass spectrometry.....	96
2.4.8.	Database searching and phosphoproteomics analyses.....	97
2.4.9.	Peptide detection, clustering and validation.....	98
2.5.	Results and Discussion.....	99
2.5.1.	Enhancement of phosphopeptide detection in LC-MS/MS experiments using FAIMS.....	99
2.5.2.	Differential phosphoproteome analysis of insulin-stimulated <i>D. melanogaster</i> cells.....	108
2.6.	Conclusion.....	117
2.7.	Acknowledgements.....	118
CHAPITRE 3 : A COMBINED PHOSPHOPROTEOMICS AND RNAI APPROACH TO IDENTIFY CK2 SUBSTRATES FOLLOWING INSULIN STIMULATION OF DROSOPHILA S2 CELLS.....		119
3.1.	Contribution des auteurs.....	120
3.2.	Abstract.....	121
3.3.	Introduction.....	122
3.4.	Materials and methods.....	124
3.4.1.	Materials.....	124
3.4.2.	Cell culture and insulin treatment.....	125
3.4.3.	RNAi synthesis and treatment.....	125
3.4.4.	Immunoblotting analysis.....	126
3.4.5.	RT-q-PCR.....	126
3.4.6.	Protein extraction.....	127
3.4.7.	Trypsin digestion.....	127
3.4.8.	TiO ₂ phosphopeptides enrichment.....	127
3.4.9.	<i>In vitro</i> kinase assays to confirm CK2 substrates.....	128
3.4.10.	Casein kinase 2 substrate peptide (CK2SP) propionylation.....	129
3.4.11.	<i>In vitro</i> kinase assay to evaluate CK2 activity.....	129
3.4.12.	Liquid chromatography.....	129
3.4.13.	FAIMS interface and mass spectrometry.....	130

3.4.14.	Database searching and phosphoproteomics analyses	131
3.4.15.	Peptide detection, clustering and validation	131
3.5.	Results and discussion	132
3.5.1.	Insulin stimulation and CK2 RNAi treatment	133
3.5.2.	Combined phosphoproteomics and RNAi analyses to identify kinase substrates regulated by insulin stimulation	134
3.5.3.	<i>In vitro</i> kinase assays confirmed CK2 direct substrates	143
3.6.	Conclusion	148
3.7.	Acknowledgements	149
 CHAPITRE 4 : OCCURRENCE AND DETECTION OF PHOSHOPEPTIDE ISOMERS IN LARGE-SCALE PHOSPHOPROTEOMICS EXPERIMENTS.....		151
4.1.	Contribution des auteurs.....	152
4.2.	Abstract	153
4.3.	Introduction.....	154
4.4.	Materials and methods	157
4.4.1.	Materials	157
4.4.2.	Cell culture and protein extraction	158
4.4.3.	Trypsin digestion.....	158
4.4.4.	TiO ₂ phosphopeptides enrichment	159
4.4.5.	Mass spectrometry	159
4.4.6.	Database search.....	160
4.4.7.	Algorithms for detecting phosphopeptide isomers from LC-MS/MS analyses.....	161
4.4.8.	Datasets availability	163
4.5.	Results and Discussion.....	164
4.5.1.	Occurrence of phosphopeptide isomers in large-scale phosphoproteomics studies	164
4.5.2.	Correlation of physical properties of phosphopeptides isomers.....	167
4.5.3.	Analysis of synthetic phosphopeptide isomers.....	169
4.5.4.	Targeted analysis of phosphopeptide isomers.....	172

4.5.5. Detection of co-eluting phosphopeptide isomers..... 174

4.5.6. Algorithm applications 178

4.6. Conclusion..... 180

4.7. Acknowledgments..... 182

CHAPITRE 5 : DISCUSSION, CONCLUSION ET PERSPECTIVES183

5.1. Discussion..... 184

5.1.1. Les défis en phosphoprotéomique 184

5.1.2. L'utilisation de LC-FAIMS-MS/MS en phosphoprotéomique 185

5.1.3. La voie de signalisation de l'insuline 187

5.1.4. L'utilisation de FAIMS pour déterminer les événements de phosphorylation associés à la voie de l'insuline chez la drosophile 187

5.1.5. Le rôle de la protéine kinase CK2 dans la voie de l'insuline 188

5.1.6. Comparaison des données phosphoprotéomiques avec la littérature disponible sur la voie de l'insuline 193

5.1.7. La détection de phosphoisomères lors d'études phosphoprotéomiques..... 197

5.2. Conclusion..... 201

5.3. Perspectives..... 203

BIBLIOGRAPHIE.....207

**ANNEXE 1 : FIGURES ET TABLEAUX SUPPLÉMENTAIRES DU CHAPITRE 1
.....XXVII**

**ANNEXE 2 : FIGURES ET TABLEAUX SUPPLÉMENTAIRES DU CHAPITRE 2
.....XXXIII**

**ANNEXE 3 : FIGURES ET TABLEAUX SUPPLÉMENTAIRES DU CHAPITRE 3
.....XXXIX**

ANNEXE 4 : FIGURES ET TABLEAUX SUPPLÉMENTAIRES DU CHAPITRE 4
.....**XLVII**

ANNEXE 5 : CONTRIBUTIONS SCIENTIFIQUES.....LVII

Liste des tableaux

Table 2-I. List of differentially phosphorylated peptides after insulin treatment in drosophila S2 cells.	112
Table 3-I. Table of statistics showing the number of identified phosphopeptides in <i>Drosophila melanogaster</i> S2 cells by LC-FAIMS-MS/MS on a LTQ-Orbitrap Velos using HCD fragmentation. .	135
Table 3-II. Examples of phosphosites affected by insulin or CK2 RNAi treatment with their corresponding intensity fold change and p-value calculated across 3 biological triplicates.....	138
Table 4-I. Types of fragment ions that reveal the presence of phosphopeptide isomers.	164
Table 4-II. Relative distribution of phosphopeptide isomers from large-scale phosphoproteomics studies.	166
Table A2- I. List of all unique phosphosites identified in FAIMS and non-FAIMS experiments conducted in control S2 cells. (CD-ROM).	xxxviii
Table A2- II. List of all unique phosphosites identified in FAIMS experiments in control and insulin-stimulated S2 cells. (CD ROM).	xxxviii
Table A3- I. List of phosphosites identified in <i>Drosophila melanogaster</i> S2 cells by 1D-LC-FAIMS-MS/MS on a LTQ-Orbitrap Velos. (CD-ROM).	xl
Table A3- II. Complete list of phosphosites affected by insulin or CK2 RNAi treatment showing their corresponding intensity fold change and p-value calculated across 3 biological triplicates. (CD-ROM).	xl
Table A4- I. List of all phosphosites identified in IEC-6, J774 and S2 cells. (CD-ROM).	lv
Table A4- II. Phosphopeptide positional isomers discovered in mouse and rat. (CD-ROM).	lv
Table A4- III. List of synthetic phosphopeptides. (CD-ROM).	lv
Table A4- IV. Phosphopeptide positional isomers in drosophila S2 cells. (CD-ROM).	lv

Liste des figures

Figure 1.1. Structures des acides aminés phosphorylables les plus fréquents (sérine, thréonine et tyrosine).....	7
Figure 1.2. Exemple de mécanisme séquentiel de phosphorylation d'un substrat par une protéine kinase.....	8
Figure 1.3. Exemples de fonctions biologiques régulées par la phosphorylation de protéines.	11
Figure 1.4. Représentation du mécanisme d'action de l'inhibition par ARN interférant chez la drosophile.....	18
Figure 1.5. Schéma global des voies de signalisation activées par l'insuline.	22
Figure 1.6. Schéma global du protocole opératoire utilisé dans cette thèse lors d'analyses phosphoprotéomiques.	31
Figure 1.7. Schéma des stratégies les plus couramment utilisées pour l'enrichissement en phosphoprotéines et phosphopeptides.	37
Figure 1.8. Spectre de masse d'un mélange de peptides représentant leur intensité en fonction de leur ratio masse sur charge (m/z).	51
Figure 1.9. Spectromètre de masse Orbitrap Elite commercialisé par Thermo Scientific.	56
Figure 1.10. Représentation de la séparation d'ions par FAIMS.....	58
Figure 1.11. Structure chimique d'un peptide et désignation des ions fragments selon la nomenclature Roepstorff–Fohlmann–Biemann suite à une fragmentation.....	63
Figure 1.12. Exemple d'un spectre MS/MS en mode CID représentant une série d'ions de type y.	64
Figure 1.13. Phénomène de β -élimination sur les résidus pSer et pThr suite à la fragmentation CID.	65
Figure 1.14. Représentation schématique de l'arbre de décision permettant de choisir en temps réel le mode de fragmentation le plus approprié selon le m/z du peptide et son état de charge.	68
Figure 1.15. Méthodologie de quantification absolue (AQUA) par spectrométrie de masse.	71
Figure 1.16. Représentation des 4 modules du logiciel Proteoprofile utilisé lors de la quantification de données protéomiques sans marquage.	84
Figure 2.1. Overview of workflow for the comparison of the phosphoproteome of S2 cells with 2D-LC-MS/MS and 2D-LC-FAIMS-MS/MS.	101
Figure 2.2. Peptides distribution in FAIMS and non-FAIMS experiments.	102
Figure 2.3. Comparison of identified phosphopeptides using FAIMS and non-FAIMS.	104
Figure 2.4. Example of phosphopeptide isomers differentiated using FAIMS.	107
Figure 2.5. Effect of insulin on the phosphorylation of ERK and Akt.....	109
Figure 2.6. Volcano plot distribution of phosphopeptides from control and insulin treated S2 cells.	111
Figure 2.7. PI3K92E (S ₁₃₆) and Sin1 (S ₁₈₄) are phosphorylated following insulin stimulation.	113
Figure 2.8. Phosphorylation motifs for phosphopeptides identified in insulin treated S2 cells.	116
Figure 3.1. Workflow of the differential phosphoproteomics experiment.	133

Figure 3.2. Phosphorylation of AKT and MAPK following insulin stimulation and CK2 knock-down after CK2 RNAi treatment.....	134
Figure 3.3. Representation of differentially phosphorylated peptides after insulin stimulation in S2 cells.....	136
Figure 3.4. Representation of differentially phosphorylated peptides after CK2 RNAi treatment in insulin-stimulated S2 cells.....	141
Figure 3.5. Examples of phosphoproteins modulated by insulin and/or CK2 RNAi treatment.....	142
Figure 3.6. Phosphorylation kinetics of synthetic peptides using <i>in vitro</i> kinase assay with CK2 recombinant enzyme.....	145
Figure 3.7. Venn diagram representation summarizing the results of the global phosphoproteomics experiment conducted in S2 cells.....	147
Figure 4.1. Separation and detection of phosphopeptide isomers.....	155
Figure 4.2. Algorithm workflow for detecting phosphopeptide isomers.....	162
Figure 4.3. Physico-chemical properties of phosphopeptide isomers separated by reverse phase LC.....	168
Figure 4.4. LC-MS separation of two isomeric phosphopeptides.....	169
Figure 4.5. Detection of synthetic phosphopeptide isomers for the peptide SLGANSSTDNVDGKK using the algorithm based on LC-MS elution profile.....	171
Figure 4.6. Identification of four phosphopeptide isomers of IPSSSSDFSK by targeted LC-MS/MS analysis.....	175
Figure 4.7. Detection of co-eluting phosphopeptide isomers using distinct fragment ion features for MS/MS spectra acquired with different activation modes.....	177
Figure A1. 1. Schéma de la stratégie PolyMAC pour l'enrichissement de phosphopeptides.....	xxviii
Figure A1. 2. Schéma de la conversion chimique d'une phosphosérine en un résidu biotinylé en vue de sa purification.....	xxix
Figure A1. 3. Séquence de six réactions chimiques menant à l'isolation sélective de phosphopeptides dans une mixture complexe par chimie de phosphoramidate.....	xxx
Figure A1. 4. Stratégies utilisant des standards isotopiques pour la quantification absolue de protéines ou de peptides par spectrométrie de masse.....	xxxï
Figure A1. 5. Schéma du protocole opératoire réalisé lors d'un marquage par TMT (Tandem Mass Tags) pour la quantification relative de peptides par spectrométrie de masse.....	xxxii

Figure A2. 1. Scatter plots of intensity measurements obtained for 2D-LC-MS/MS experiments performed with and without FAIMS.	xxxiv
Figure A2. 2. Scatter plot of intensity measurements for peptides identified in both FAIMS and non-FAIMS 2D-LC-MS/MS experiments.	xxxv
Figure A2. 3. Variation of CV values for phosphopeptides isomers.	xxxvi
Figure A2. 4. Scatter plot of intensity measurements obtained for 2D-LC-MS/MS analyses of control and insulin treated <i>D. melanogaster</i> S2 cells (8 min stimulation, 0.86 μM insulin).	xxxvii
Figure A4. 1. Physico-chemical properties of phosphopeptide isomers separated by reverse phase LC.	xlvi
Figure A4. 2. The order of elution of phosphopeptide isomers is related to proximity of the phosphorylated residue to the N-terminus.	xlix
Figure A4. 3. Overlap in phosphopeptide identification between CID, HCD and ETD for <i>D. melanogaster</i> S2 phosphoproteome.	l
Figure A4. 4. Difference of retention time for 117 phosphopeptide isomers of <i>D. melanogaster</i> S2 cells using DDA (survey) and targeted MS/MS analyses.	li
Figure A4. 5. Properties of 117 phosphopeptide isomers of <i>D. melanogaster</i> S2 cells.	lii
Figure A4. 6. Isobaric peptide artifacts.	lii
Figure A4. 7. LC separation of phosphopeptide conformers.	liii
Figure A4. 8. MS/MS spectra of four phosphopeptide isomers from IPSSSSDFSK.	liv

Liste des abréviations

4E-BP	<i>EIF4E-binding protein</i>
a.a.	acide aminé
ACN	Acétonitrile
ADN	Acide désoxyribonucléique
AGC	Gain de contrôle automatique (<i>Automatic Gain Control</i>)
AIMS	Spectrométrie par mobilité ionique d'aspiration (<i>Aspiration Ion Mobility Spectrometry</i>)
AKH	Hormone adipocinétique (<i>Adipokinetic Hormone</i>)
AMP	Adénosine monophosphate
amu	unité de masse atomique (<i>atomic mass unit</i>)
APCI	Ionisation chimique à pression atmosphérique (<i>Atmospheric-Pressure Chemical Ionization</i>)
APS	<i>Adapter protein with Pleckstrin homology and Src homology 2 domains</i>
AQUA	Quantification absolue (<i>Absolute Quantification</i>)
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messager
ATM	<i>Ataxia telangiectasia-mutated</i>
ATP	Adénosine triphosphate
ATR	<i>ATM and Rad3-related</i>
blast	outil de recherche d'alignement basique local (<i>basic local alignment search tool</i>)
BRCA1	<i>Breast Cancer type 1 susceptibility protein</i>
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CAD	Dissociation activée par collision (<i>Collision-Activated Dissociation</i>)
CAP	<i>adenylyl Cyclase-Associated Protein 1</i>
CBL	<i>Casitas B-Lineage lymphoma proto-oncogene</i>
CDC2	<i>Cell Division Control protein 2</i>

CDC25	<i>Cell Division Cycle 25</i>
CDPK	Calcium-Dependent Protein Kinase
CEACAM1	<i>Carcinoembryonic Antigen-related Cell Adhesion Molecule</i>
CID	Dissociation induite par collision (<i>Collision-Induced Dissociation</i>)
CK1	Caséine kinase 1
CK1SP	Peptide substrat de CK1 (<i>Casein kinase 1 substrate peptide</i>)
CK2	Caséine kinase 2
CK2SP	Peptide substrat de CK2 (<i>Casein kinase 2 substrate peptide</i>)
.csv	<i>comma-separated values</i>
C-ter	Extrémité carboxyl-terminale
CV	Voltage de compensation (<i>Compensation Voltage</i>)
Da	Dalton
DDA	Acquisition dépendante des données (<i>Data-dependent acquisition</i>)
DESI	Ionisation par désorption électro-ébulisaison (<i>Desorption Electrospray Ionization</i>)
DHB	Acide 2,5-dihydroxybenzoïque
DOCK1	<i>Dedicator Of Cytokinesis protein 1</i>
DOCK2	<i>Dedicator Of Cytokinesis protein 2</i>
dsRNA	ARN double brin (<i>double strand RNA</i>)
DTIMS	Spectrométrie par mobilité ionique en fonction du temps de dérive (<i>Drift-Time Ion Mobility Spectrometry</i>)
DTT	Dithiothréitol
DV	Voltage de dispersion (<i>Dispersion Voltage</i>)
EDTA	Acide éthylène diamine tétraacétique
EGF	Facteur de croissance épidermal (<i>Epidermal Growth Factor</i>)
EGFR	Récepteur du facteur de croissance épidermal (<i>Epidermal Growth Factor Receptor</i>)
EGTA	Acide éthylène glycol tétraacétique
EIF2B	<i>Eukaryotic Initiation Factor 2B</i>

EIF4E	<i>Eukaryotic Initiation Factor 4E</i>
EPHB2	<i>Ephrin type-B receptor 2</i>
ERK	<i>Extracellular signal-regulated kinase</i>
ESI	<i>Électronébulisation (Electrospray Ionisation)</i>
ETD	<i>Dissociation par transfert d'électron (Electron Transfer Dissociation)</i>
ExPASy	<i>Expert Protein Analysis System</i>
FA	<i>Acide formique</i>
FAB	<i>Bombardement par des atomes rapides (Fast Atom Bombardment)</i>
FAIMS	<i>Spectrométrie de mobilité ionique modulée par un champ électrique en forme d'onde asymétrique à champ élevé (high-Field Asymmetric waveform Ion Mobility Spectrometry)</i>
FBS	<i>Sérum de veau fœtal (Fetal Bovine Serum)</i>
FC	<i>Fold change</i>
FDR	<i>Taux de faux positifs (False Discovery Rate)</i>
fmol	<i>femtomol</i>
FOXO	<i>Forkhead box protein O</i>
FT-ICR	<i>Résonance ionique cyclotronique et transformée de Fourier (Fourier Transform Ion Cyclotron Resonance)</i>
GAB1	<i>Growth factor receptor bound protein 2-associated protein 1</i>
GAB2	<i>Growth factor receptor bound protein 2-associated protein 2</i>
GAP	<i>GTPase-Activating Protein</i>
GFP	<i>Green Fluorescent Protein</i>
GLUT4	<i>Glucose Transporter Type 4</i>
GO	<i>Gene Ontology</i>
GRB2	<i>Growth factor Receptor-Bound protein 2</i>
GSK3	<i>Glycogen Synthase Kinase 3</i>
GST	<i>Glutathione S-transférase</i>
GTP	<i>Guanosine triphosphate</i>
HAD	<i>Haloacid Dehalogenase</i>

HCD	Dissociation par collision de plus haute énergie (<i>Higher-energy Collisional Dissociation</i>)
HPLC	Chromatographie liquide haute performance (<i>High Performance Liquid Chromatography</i>)
ICAT	Étiquette d'affinité codée par un isotope (<i>Isotope-Coded Affinity Tag</i>)
i.d.	diamètre interne (<i>internal diameter</i>)
IDA	Acide iminodiacétique
IEC-6	<i>Intestinal Epithelial Cell line, No. 6</i>
IGF	<i>Insulin-like Growth Factor</i>
Iip	<i>Insulin-like protein</i>
IMAC	Chromatographie d'affinité au métal immobilisé (<i>Immobilized Metal ion Affinity Chromatography</i>)
IMS	Spectrométrie par mobilité ionique (<i>Ion Mobility Spectrometry</i>)
InR	<i>Insulin-like Receptor</i>
IPC	Cellule produisant de l'insuline (<i>Insulin Producing Cell</i>)
IPI	Index de protéine international (<i>International Protein Index</i>)
IRF4	<i>Interferon Regulatory Factor 4</i>
IRS-1	<i>Insulin Receptor Substrate 1</i>
IRS-2	<i>Insulin Receptor Substrate 2</i>
IRS-3	<i>Insulin Receptor Substrate 3</i>
IRS-4	<i>Insulin Receptor Substrate 4</i>
iTRAQ	Étiquette isobarique pour la quantification relative et absolue de peptide (<i>Isobaric peptide Tag for Relative and Absolute Quantification</i>)
IVKA	Essai kinase <i>in vitro</i> (<i>In Vitro Kinase Assay</i>)
JIP	<i>JNK-Interacting Protein</i>
JNK	<i>c-Jun N-terminal Kinase</i>
LC-MS/MS	Chromatographie liquide couplée à la spectrométrie de masse en tandem
LIT	Trappe ionique linéaire (<i>Linear Ion Trap</i>)
LTQ	Trappe linéaire quadrupolaire (<i>Linear Trap Quadrupole</i>)

Lys-C	Lysyl endoprotéinase
MALDI	Désorption/ionisation assistée par matrice (<i>Matrix-Assisted Laser Desorption/Ionization</i>)
MAPK	<i>Mitogen-Activated Protein Kinase</i>
MAPK-APK	<i>MAP Kinase-Activated Protein Kinase</i>
MAPKK	<i>MAP Kinase Kinase</i>
MAX	<i>Myc-Associated factor X</i>
MDa	Million de Daltons
MEK	<i>Map kinase/Erk Kinase</i>
MeOH	Méthanol
.mgf	<i>mascot generic file</i>
MKK7	<i>Map Kinase Kinase 7</i>
MKP	<i>Map Kinase Phosphatase</i>
MS	Spectrométrie de masse (<i>Mass Spectrometry</i>)
MS/MS	Spectrométrie de masse en tandem
MudPIT	Technologie d'identification de protéine en plusieurs dimensions (<i>Multidimensional Protein Identification Technology</i>)
MUNC18	<i>Mammalian Uncoordinated-18</i>
m/z	ratio masse sur charge
NES	Signal d'export nucléaire (<i>Nuclear Export Signal</i>)
NLS	Signal de localisation nucléaire (<i>Nuclear Localization Signal</i>)
NTA	Acide nitrilotriacétique
N-ter	Extrémité NH ₂ -terminale
PAC	Chimie de phosphoramidate (<i>Phosphoramidate Chemistry</i>)
PBS	Tampon phosphate salin (<i>Phosphate Buffered Saline</i>)
PDB	<i>Protein Data Bank</i>
PDGFR	Récepteur du facteur de croissance dérivé des plaquettes (<i>Platelet-Derived Growth Factor Receptor</i>)
PDK1	<i>Phosphoinositide-Dependent Kinase 1</i>

PDMS	Spectrométrie de masse de désorption au plasma (<i>Plasma desorption mass spectrometry</i>)
PI(3,4,5)P3	Phosphoinositide 3,4,5 triphosphate
PI3K	Phosphoinositide-3 Kinase
PKA	Protéine Kinase A
PKC	Protéine Kinase C
PMF	Empreinte peptidique massique (<i>Peptide Mass Fingerprint</i>)
PolyMAC	Capture par affinité aux ions métalliques basés sur un polymère (<i>Polymer-based Metal ion Affinity Capture</i>)
PP1	<i>Protein Phosphatase 1</i>
PP2A	<i>Protein Phosphatase 2A</i>
PP2C	<i>Protein Phosphatase 2C</i>
ppm	partie par million
PPM	Protéines phosphatases dépendantes du Mg ²⁺ ou du Mn ²⁺
PPP	Phosphoprotéine phosphatase
pSer	Phosphosérine
PSAQ	Quantification absolue par standard de protéine (<i>Protein Standard Absolute Quantification</i>)
pThr	Phosphothréonine
PTB	Liaison aux phosphotyrosines (<i>Phosphotyrosines binding</i>)
PTEN	<i>Phosphatase and Tensin homolog</i>
PTM	Modification post-traductionnelle (<i>Post-Translational Modification</i>)
PTP	Protéine tyrosine phosphatase
pY	Phosphotyrosine
QconCAT	Quantification par concatamère
RF	Fréquence radio (<i>Radio Frequency</i>)
RHEB	<i>Ras-Homologue Expressed in Brain</i>
Rho-GEF	<i>Rho-family Guanine nucleotide Exchange Factor</i>
RISC	<i>RNA-Induced Silencing Complex</i>

RNAi	ARN interférant
RP-HPLC	Chromatographie liquide haute performance sur phase inverse (<i>Reversed Phase High-Performance Liquid Chromatography</i>)
RSD	Déviatoin standard relative (<i>Relative Standard Deviation</i>)
rt	temps de rétention (<i>retention time</i>)
RTK	Récepteur tyrosine kinase
S6K	<i>Ribosomal protein S6 Kinase</i>
SAX	Échange fort d'anions (<i>Strong Anion Exchange</i>)
SCX	Échange fort de cations (<i>Strong Cation Exchange</i>)
SDS-PAGE	Électrophorèse sur gel de polyacrylamide avec dodécyl sulfate de sodium (<i>Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis</i>)
SH2	<i>Src Homology 2</i>
SHC	<i>Src Homology 2 domain Containing</i>
SILAC	Marquage isotopique stable en culture cellulaire (<i>Stable Isotope Labeling by Amino acids in Cell culture</i>)
SIMAC	Élution séquentielle suivant IMAC (<i>Sequential elution from IMAC</i>)
SIMS	Spectrométrie de masse d'ion secondaire (<i>Secondary-Ion Mass Spectrometry</i>)
siRNA	petit ARN interférant (<i>small interfering RNA</i>)
SNARE	<i>SNAP (Soluble NSF Attachment Protein) Receptor</i>
SOS	<i>Son Of Sevenless</i>
SRM	<i>Selected Reaction Monitoring</i>
STAT-1	<i>Signal Transducer and Activator of Transcription 1</i>
TAP	Purification par affinité en tandem (<i>Tandem Affinity Purification</i>)
TCEP	Tris[2-carboxyethyl] phosphine
TFA	Acide trifluoroacétique
TiO ₂	Dioxyde de titane
TMT	Étiquettes de masse en tandem (<i>Tandem Mass Tags</i>)
TOP	Topoisomérase

TOF	Tube de temps de vol (<i>Time Of Flight</i>)
TOR	<i>Target Of Rapamycin</i>
TOR-C1	<i>TOR Complex 1</i>
TSC1	<i>Tuberous Sclerosis Complex 1</i>
TSC2	<i>Tuberous Sclerosis Complex 1</i>
TWIMS	Spectrométrie de mobilité ionique par onde progressive (<i>Traveling-Wave Ion Mobility Spectrometry</i>)
UV	Ultraviolet
V	Volt
VAMP2	<i>Vesicle-Associated Membrane Protein 2</i>
VEGFR	Récepteur du facteur de croissance épidermal vasculaire (<i>Vascular Endothelial Growth Factor Receptor</i>)
XIC	Courant ionique extrait (<i>Extracted Ion Current</i>)

Remerciements

En premier lieu, je remercie mon directeur de recherche **Dr. Pierre Thibault** pour m'avoir accueillie dans son laboratoire. Ce fut un plaisir de faire partie de votre groupe de recherche et d'interagir avec vous au quotidien. Grâce à votre appui, j'ai pu bénéficier d'une formation hors-norme dans le domaine de la protéomique en ayant accès à des instruments fiables et à la fine pointe de la technologie (FAIMS).

Je voudrais remercier plus particulièrement Chantal Durette, Louiza Mahrouche et Christelle Pomiès avec qui je suis devenue amie tant au laboratoire qu'à l'extérieur. Un gros merci au Dr. Éric Bonneil pour son soutien inestimable au niveau instrumental et pour avoir partagé son expérience lors de l'écriture des publications. Merci à Mathieu Courcelles pour son aide précieuse et continue dans le domaine bioinformatique et pour notre ultime collaboration lors de notre publication commune. J'aimerais souligner la contribution d'Olivier Caron-Lizotte pour la grande patience dont il fait preuve quand il s'agit de trouver une solution informatique à divers problèmes de quantification. De plus, son allergie aux œufs nous assurait à tous une part de gâteau supplémentaire les jours de fête. Un merci spécial à toutes les personnes avec qui j'ai travaillé au laboratoire : Alexandre, Antoine, Christina, Danielle, Dev, Émilie, Evgeny, Fabyana, Frédéric, Julian, Kevin, Maria, Marie-Hélène, Marjo, Marlène, Matthias, Michel, Nebiyu, Olivier, Paul, Raquel, Tara et Wafaa.

Je voudrais remercier Alain Verreault pour son aide précieuse, son écoute attentionnée et pour les discussions que nous avons partagées. Les cours d'été m'ont beaucoup appris et ce fut un plaisir d'enseigner avec toi.

Je remercie les membres de mon comité de thèse (Dr. Vincent Archambault et Dr. Sylvain Meloche) pour leur avis pertinent au cours de nos rencontres annuelles ainsi que le jury de mon examen prédoctoral (Dr. Marc Therrien, Dr. Sébastien Carréno et Dr. Stephen Michnick).

Je voudrais citer les membres du laboratoire d'enseignement du département de biochimie (Daniel, Nathalie, André, Nicole, Philippe et Shona) avec lesquels j'ai passé de

très bon moments pendant la durée des travaux pratiques sur la mitochondrie. Ce fut une expérience très enrichissante (pourquoi certains étudiants trouvent ça drôle de se mettre du cyanure dans le nez à 20h30 le soir ?). Ce fut inoubliable.

Je voudrais remercier ma famille pour m'avoir encouragée tout au long de mon parcours scolaire et supportée lors de ma décision de partir à l'étranger.

Je remercie Gaëtan qui m'a soutenue et écoutée ~~me plaindre~~ tout au long de mon doctorat, tu es bien courageux !

Merci !

Chapitre 1 : Introduction

1.1. Organisation des chapitres

Le **Chapitre 1 : Introduction** décrit le rôle global de la phosphorylation dans la signalisation cellulaire ainsi que les méthodes utilisées dans cette thèse pour son analyse. Les implications biologiques et biochimiques présentées dans cette section regroupent la voie de signalisation de l'insuline et l'utilisation de la drosophile comme organisme modèle. Les différentes techniques de préparation d'échantillon en vue d'analyses phosphoprotéomiques sont abordées, incluant notamment les méthodes d'enrichissement en phosphopeptides. Les aspects instrumentaux tels que la spectrométrie de masse (MS), la spectrométrie de mobilité ionique modulée par un champ électrique en forme d'onde asymétrique à champ élevé (FAIMS) ainsi que les méthodes de séparation chromatographiques sont détaillés. Les outils d'analyses bioinformatiques utilisés sont mentionnés pour l'étude des aspects qualitatif et quantitatif des données. La description des objectifs du projet est présentée à la fin de ce chapitre.

Le **Chapitre 2 : Improvement of phosphoproteome analyses using FAIMS and decision tree fragmentation; Application to the insulin signalling pathway in *Drosophila melanogaster* S2 cells** présente mon premier article en tant que co-premier auteur publié dans *Journal of Proteome Research* (2012)[1]. Cette étude a pour but d'évaluer les bénéfices de l'utilisation du FAIMS en combinaison avec la séparation chromatographique liquide (LC) et la spectrométrie de masse dans le contexte d'analyses phosphoprotéomiques. Nous décrivons la capacité de FAIMS à séparer les peptides multiplement chargés comparativement aux ions contaminants interférants (solvant, phase stationnaire...) afin d'améliorer la détection de phosphopeptides de faible abondance. Nous observons une augmentation significative du ratio signal sur bruit, permettant ainsi d'accroître de 50 % le nombre de phosphopeptides identifiés suite à l'utilisation du FAIMS comparativement à une étude conventionnelle. Nous avons utilisé cette nouvelle stratégie afin de disséquer les événements de phosphorylation différentiels en réponse au traitement à l'insuline dans des cellules S2 de drosophile.

Le Chapitre 3 : A combined phosphoproteomics and RNAi approach to identify CK2 substrates following insulin stimulation of Drosophila S2 cells présente mon deuxième article en tant que premier auteur (manuscrit en préparation) en voie de soumission à *Journal of Proteome Research*. Une nouvelle stratégie expérimentale combinant l'utilisation de LC-FAIMS-MS/MS et ARN interférant (RNAi) est utilisée afin d'identifier les événements de phosphorylation dépendant de la stimulation à l'insuline et de la caséine kinase 2 (CK2) dans des cellules S2 de drosophile. Nous avons utilisé des logiciels bioinformatiques qui permettent l'identification des kinases susceptibles de phosphoryler un site d'après sa séquence consensus ainsi que les informations disponibles dans la littérature. Ceci a révélé qu'environ 50 % des sites régulés par l'insuline se situent dans un environnement acide constitué d'acides aminés ciblés par CK2. Une étude phosphoprotéomique quantitative sans marquage a été réalisée et a mis en évidence plusieurs cibles d'intérêt dont la phosphorylation est dépendante à la fois de l'insuline et de CK2. Des essais kinase *in vitro* ont permis de confirmer les candidats qui sont des cibles directes de CK2.

Le Chapitre 4 : Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments présente mon troisième article en tant que co-premier auteur publié dans *Journal of Proteome Research* (2012)[2]. Dans cet article, nous avons développé deux outils bioinformatiques afin d'améliorer l'identification de phosphopeptides isomériques (même séquence peptidique mais position alternative du groupement phosphate) lors d'études phosphoprotéomiques. En effet, peu d'information est disponible dans la littérature quant à la proportion réelle des phosphoisomères positionnels. Le premier algorithme permet de distinguer les isoformes séparables par LC-MS en se basant sur leur différence de temps de rétention. La seconde stratégie consiste à différencier les formes qui co-éluent ou qui sont partiellement séparées en étudiant la présence de fragments caractéristiques qui vont confirmer la position du phosphate sur un résidu précis par MS/MS. Dans les deux cas, des listes d'inclusion basées sur les résultats des algorithmes sont générées afin de valider les candidats lors d'une seconde expérience ciblée. Ces études complémentaires ont montré qu'environ 1 % des phosphopeptides

identifiés sont isomériques. La confiance de localisation du site modifié s'avère cruciale dans le cas où une fonction biologique associée à cette position veut être étudiée plus en détail.

1.2. Les protéines

Les protéines sont des composés biochimiques présents dans la cellule au même titre que les lipides, les glucides ou les acides nucléiques. Elles jouent un rôle essentiel dans de nombreux processus biologiques et participent également à la structure de la cellule.

Elles sont générées suite à la transcription de gènes contenus dans l'ADN (acide désoxyribonucléique) pour former l'ARNm (acide ribonucléique messenger) qui sert de modèle pour la synthèse protéique suite à sa traduction par les ribosomes. Ce sont les séquences spécifiques en nucléotides des gènes qui déterminent les séquences protéiques. Elles sont constituées d'acides aminés (20 molécules standards) liés entre eux par des liens peptidiques covalents qui relient les extrémités C-terminale (C-ter) et N-terminale (N-ter) de deux résidus adjacents.

Les protéines possèdent quatre ordres de structures hiérarchiques essentiels à leur fonction soit la structure primaire (séquence en acides aminés), secondaire (structure locale telle que les hélices α ou feuillets β), tertiaire (repliement spatial entre les structures secondaires) et quaternaire (interactions entre protéines). Elles peuvent se regrouper sous forme de complexe afin de réguler divers processus. Ce sont les principaux acteurs permettant de véhiculer l'information codée par le matériel génétique et induire une transduction de signal dans la cellule. On pense qu'il existe entre 20 000 et 40 000 protéines chez l'humain (selon les cadres ouverts de lecture prédits) mais les isoformes générés suite à l'épissage alternatif ou à différents clivages enzymatiques sous-estiment ces prédictions.

Suite à leur synthèse, les protéines peuvent être modifiées de manière chimique (phosphorylation, acétylation, glycosylation, ubiquitylation...), ce qui influence leur stabilité, leur repliement, leur activité ainsi que leur fonction. Dans cette thèse, nous nous sommes concentrés principalement sur l'étude de la phosphorylation.

1.3. La phosphorylation

1.3.1. Mécanisme général

La phosphorylation des protéines est une modification post-traductionnelle (PTM) importante qui régule de nombreux processus cellulaires et qui est impliquée dans diverses voies de signalisation. Elle a été mise en évidence pour la première fois par P. A. Levene et C. L. Alsberg en 1906[3]. On considère qu'environ 30 % des protéines sont susceptibles d'être phosphorylées à un certain moment[4, 5] et on estime que plus de 100 000 sites de phosphorylation existent chez l'humain[4]. La phosphorylation joue un rôle capital et contrôle divers processus incluant la transcription, la traduction, l'apoptose, le réarrangement du cytosquelette ou encore le métabolisme. C'est une modification réversible qui permet de réguler l'activité d'une protéine en changeant sa conformation et en modulant ses interactions avec d'autres protéines.

La phosphorylation consiste en l'addition d'un groupement phosphate (PO_4^{3-}) principalement sur les résidus sérine (80 %), thréonine (17 %) et tyrosine (3 %) (**Figure 1.1**). Bien que moins fréquente, la phosphorylation des tyrosines reste néanmoins d'une importance biologique essentielle. La phosphorylation peut également se produire sur les histidines ou aspartates en moindre mesure. Chez les procaryotes, la phosphorylation des arginines et lysines est aussi observée.

1.3.2. Kinases et phosphatases

Les protéines kinases sont les enzymes responsables de la phosphorylation et agissent de manière coordonnée avec les phosphatases impliquées dans la déphosphorylation réciproque.

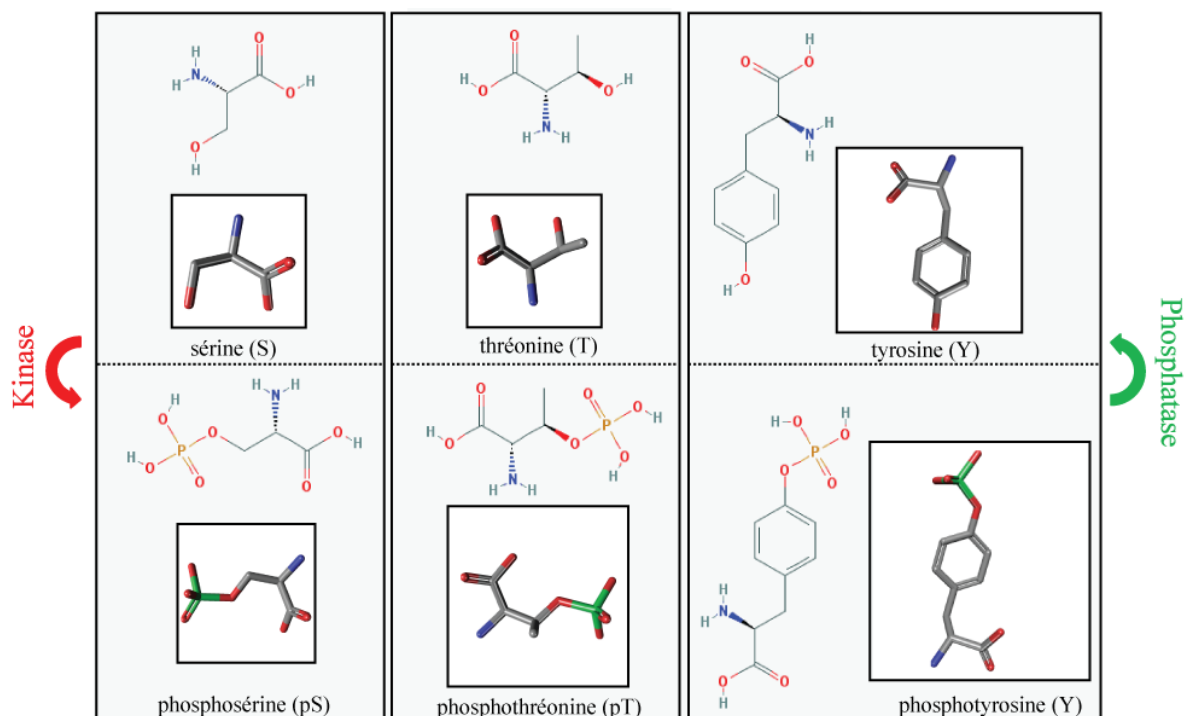


Figure 1.1. Structures des acides aminés phosphorylables les plus fréquents (sérine, thréonine et tyrosine).

La phosphorylation et déphosphorylation des acides aminés sont respectivement effectuées par les protéines kinases et phosphatases. Les structures sont obtenues à partir de la base de données PubChem Compound.

Les kinases reconnaissent leur substrat et catalysent le transfert d'un groupe phosphate d'un nucléoside triphosphate (principalement l'ATP [adénosine triphosphate] ou le GTP [guanosine triphosphate]) sur le groupement hydroxyle (-OH) libre des chaînes latérales des acides aminés porteurs d'une fonction alcool (**Figure 1.2**). Le groupe -OH neutre est alors remplacé par un phosphate porteur d'une charge négative attaché de manière covalente. La plupart des kinases agissent à la fois sur les sérines et thréonines tandis que d'autres sont spécifiques aux tyrosines. On trouve également des kinases ayant une spécificité mixte qui sont capables de phosphoryler les 3 types de résidus[6]. Les kinases peuvent également s'autophosphoryler. Il existe 27 familles de kinases qui représentent environ 2 % des gènes chez les eucaryotes où 2 000 kinases ont été recensées. Chez l'humain, une étude a rapporté 518 gènes codant pour des kinases potentielles, 540 chez la souris, 251 chez la drosophile

et 123 chez la levure[7]. Il faut mentionner que d'une manière générale, la sous-unité catalytique (comprenant un acide aspartique) des protéines kinases est très conservée d'un point de vue structural. Le domaine liant l'ATP est quant à lui généralement riche en glycines et contient une lysine. Ainsi, de nombreux inhibiteurs ont une action dirigée vers ces enzymes lors de l'élaboration de traitements ciblés pour le cancer par exemple.

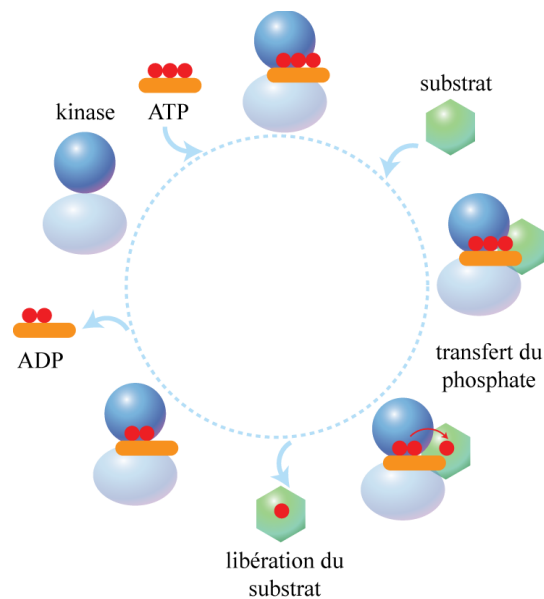


Figure 1.2. Exemple de mécanisme séquentiel de phosphorylation d'un substrat par une protéine kinase.

L'ATP se lie au site catalytique de la kinase et induit la liaison du substrat au site actif. Par la suite, le phosphate γ de l'ATP est transféré sur un résidu phosphorylable du substrat. Après sa phosphorylation, le substrat est libéré de la kinase et l'ADP est relâché du site actif. L'ordre des étapes peut varier selon le mode d'action de la kinase (par exemple, certaines kinases lient leur substrat avant l'ATP ou d'autres libèrent l'ADP avant le substrat).

Les protéines sérine ou thréonine kinases sont régulées par diverses molécules telles que des nucléotides cycliques (l'AMP [adénosine monophosphate] cyclique pour PKA [protéine kinase A]), le diacylglycérol (pour PKC [protéine kinase C]) ou encore le calcium (pour CDPK [calcium-dependent protein kinase] chez les plantes et les protozoaires). Les protéines tyrosines kinases sont majoritairement cytoplasmiques mais les récepteurs tyrosine kinases (RTKs) appartiennent à une classe bien particulière et consistent en un récepteur transmembranaire (généralement sous la forme d'une hélice α) contenant un domaine tyrosine kinase dirigé vers le cytoplasme. Deux types de RTKs se distinguent par leur mécanisme d'activation. Certains sont préassemblés de manière constitutive sous forme de dimère, ce qui est le cas par exemple du récepteur de l'insuline (InR) dont les deux sous-unités sont liées par un pont disulfure. La liaison de l'insuline est cependant requise pour l'activation du RTK suite à un changement de conformation permettant l'autophosphorylation des deux sous-unités. Pour d'autres RTKs, la liaison du ligand (petite molécule, protéine...) change leur structure quaternaire et induit leur dimérisation. Généralement, la liaison entre le ligand et le récepteur se fait de manière coopérative. Le ligand se lie d'abord à une des deux sous-unités monomériques, ce qui favorise son association avec la seconde sous-unité du récepteur. La dimérisation est cruciale afin que les domaines kinases puissent s'autophosphoryler au niveau d'un segment d'activation, ce qui induit le déplacement des deux sous-domaines de la kinase intrinsèque. Ceci provoque l'ouverture du domaine kinase libre pour la liaison de l'ATP[8]. C'est notamment le cas du récepteur au facteur de croissance épidermal (EGF) ou encore du récepteur du facteur de croissance dérivé des plaquettes (PDGFR) qui se trouvent sous forme de monomère en absence d'hormone. Ce mécanisme bien connu régule l'activité kinase en passant d'un état non-actif à un état « en attente » prêt à induire une transduction de signal.

Chaque kinase a une certaine spécificité vis-à-vis de son substrat afin de communiquer un signal précis. En se basant sur des analyses statistiques portant sur un grand nombre de séquences substrats de kinases, il a été démontré que les 4 acides aminés en amont et en aval du site accepteur de phosphate ont une influence sur la spécificité et la sélection du substrat. En effet, on peut classer globalement les séquences substrats en 3

grandes catégories : les acidophiles (caséine kinase), basophiles (PKA, PKC) et celles dirigées par la proline (MAPK [mitogen-activated protein kinase], P38, kinases cyclines dépendantes). Ainsi, les kinases reconnaissent spécifiquement des motifs constitués d'acides aminés caractéristiques qui forment une séquence consensus cible[9]. Certains motifs sont partagés par plusieurs kinases comme le motif S/T-P commun aux interacteurs impliqués dans la voie MAPK. Des événements de phosphorylation multiples sont possibles au sein d'un même site consensus dans le cas de boucles d'activation par exemple.

Les protéines phosphatases ont pour fonction d'enlever le groupement phosphate de leur substrat en hydrolysant le monoester d'acide phosphorique en un ion phosphate, ce qui laisse le résidu avec un groupement hydroxyle libre. En se basant sur leur homologie de séquence et leur structure tridimensionnelle, les phosphatases sont classées en quatre groupes distincts : les phosphoprotéines phosphatases PPP (PP1, PP2A...), les protéines phosphatases PPM dépendantes du Mg^{2+} ou du Mn^{2+} (PP2C...), les protéines tyrosine phosphatases PTP (CDC25...) et les protéines phosphatases spécifiques à l'aspartate (HAD...)[10]. Plus de 100 phosphatases ont été prédites par annotation du génome chez l'humain[11].

La phosphorylation ou déphosphorylation d'une protéine ne correspond pas respectivement à son activation ou sa désactivation puisque chacun des résidus modifiés influence la fonction globale de la protéine. En effet, une protéine peut être multiplement phosphorylée sur des sites distincts qui correspondent à des sites d'activation ou de répression. La synergie globale résultant de toutes les modifications peut influencer l'activité de la protéine de manière directe ou en modifiant sa localisation cellulaire au sein de la cellule (comme dans le cas de facteurs de transcription). Par exemple, STAT-1 est transloqué au noyau en réponse à un traitement à l'interféron-gamma et induit la régulation de gènes. La phosphorylation peut également moduler des interactions protéine-protéine ou encore la dégradation par le protéasome en coordination avec l'ubiquitylation (**Figure 1.3**).

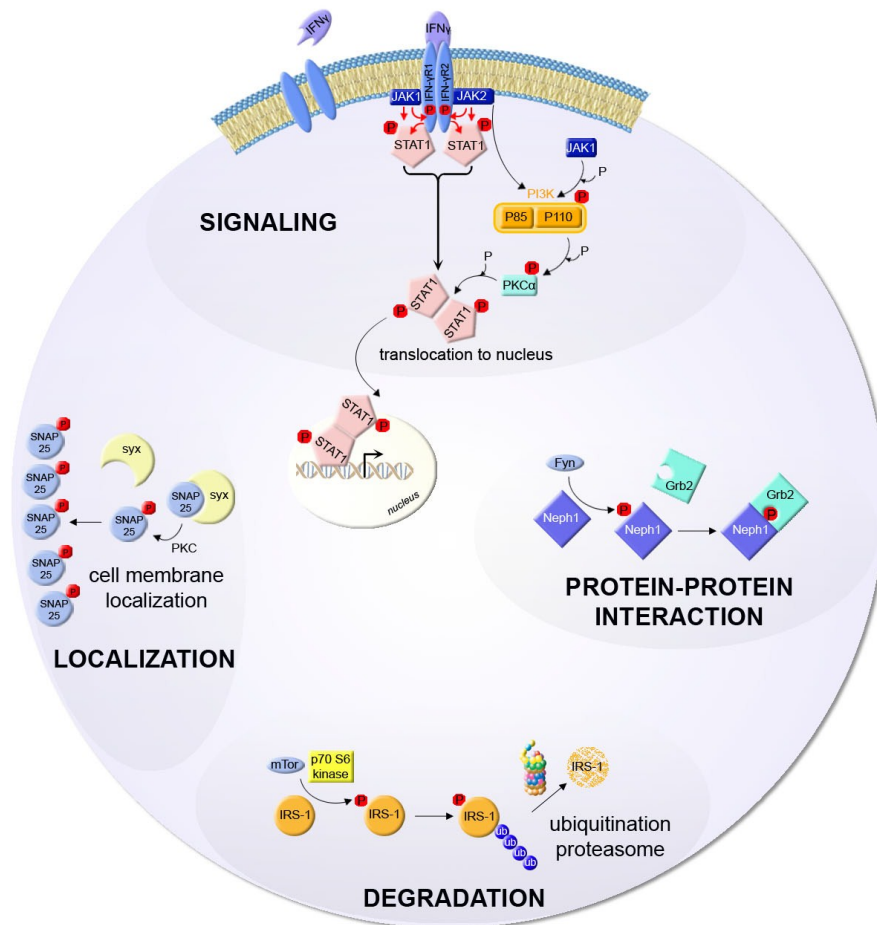


Figure 1.3. Exemples de fonctions biologiques régulées par la phosphorylation de protéines.

Signalisation : la liaison de l'interféron-gamma à son récepteur induit son autophosphorylation et engendre la formation d'un complexe formé par JAK1 et JAK2. Ceci occasionne la phosphorylation de STAT-1 et sa translocation au noyau où il va réguler la transcription de gènes. **Localisation** : la phosphorylation de SNAP25 réduit son affinité de liaison avec la syntaxin-1A et induit sa translocation vers la membrane plasmique. **Interaction protéine-protéine** : la phosphorylation de NEPH1 par FYN favorise son interaction avec GRB2. **Dégradation** : la phosphorylation du substrat du récepteur de l'insuline (IRS-1) par la protéine ribosomale S6 kinase β -1 induit sa polyubiquitylation par CUL-7 et sa dégradation par le protéasome. Cité de [12] avec la permission de John Wiley and Sons, Mass Spectrometry Reviews, copyright 2010.

Cependant, il faut noter que la phosphorylation peut aussi n'avoir aucune conséquence biologique et aucune fonction associée[13]. Il est raisonnable de penser qu'un site de phosphorylation inutile peut persister au cours de l'évolution si aucun désavantage concret ne lui est attribué. Il a été rapporté que les sites ayant une fonction importante sont plus conservés entre espèces par rapport aux sites provenant d'effets « hors-cibles » générés par certaines kinases au hasard. Ceci engendre des événements de phosphorylation de type « bruit » qui se produisent le plus souvent dans des régions désordonnées où les sites sont plus accessibles. A ce jour, PhosphositePlus[14] (<http://www.phosphosite.org>) rapporte 166 758 sites de phosphorylation non redondants sur 18 688 protéines chez l'humain et la souris. La majorité de ces sites n'a aucune fonction attribuée jusqu'à présent, soit parce qu'elle n'est pas encore connue ou bien parce qu'elle n'a aucun rôle significatif *in vivo*. Cependant, leur proportion respective est encore mystérieuse.

1.3.3. La phosphoprotéomique

La phosphoprotéomique consiste en l'étude du phosphoprotéome, soit l'ensemble des phosphoprotéines exprimées à un moment précis et dans un environnement donné (organe, tissu, cellule, organelle...). Un génome va conduire à différents protéomes dépendamment du stade de différenciation cellulaire, d'une réponse à un signal ou encore d'un état pathologique ce qui va également avoir des répercussions au niveau traductionnel et post-traductionnel. La phosphoprotéomique est une sous-catégorie de la protéomique (au même titre que la glycoprotéomique) basée sur l'identification globale des événements de phosphorylation qui inclut également leur quantification suite à une stimulation cellulaire par exemple. Combinées à la spectrométrie de masse, les études phosphoprotéomiques constituent une approche puissante dans le but d'élucider des mécanismes impliqués dans diverses voies de signalisation et de quantifier les changements dynamiques de phosphorylation de manière cinétique.

1.3.4. L'intérêt clinique de l'étude de la phosphorylation

Les analyses phosphoprotéomiques sont d'un intérêt capital étant donné que les modulations de phosphorylation ont une importance prédominante dans le développement de nombreux types de cancers[15].

Au Canada, on estime que 177 800 nouveaux cas de cancer surviendront en 2012 et 45 % des Canadiens en seront atteints au cours de leur vie (source : Société Canadienne du Cancer). D'après les statistiques, un Canadien sur quatre meurt du cancer. Il s'agit d'une pathologie caractérisée par la présence de tumeurs malignes formées à partir de la transformation par mutations (ou instabilité génétique) d'une cellule initialement normale. Il est défini par une prolifération cellulaire anormale et illimitée de telle manière que la survie des organes est menacée. La transformation cellulaire tumorale se traduit notamment par une perte de contrôle du cycle cellulaire, une insensibilité à l'apoptose et des anomalies au niveau de la réparation de l'ADN. Certains facteurs peuvent expliquer son apparition comme la présence d'un agent infectieux (virus, bactérie ou parasite), des facteurs externes ou environnementaux (alimentation, exposition à des composés cancérigènes...), des causes hormonales ou encore épigénétiques. La maladie peut évoluer suite à la formation de métastases qui consistent en des cellules qui migrent de leur lieu de production et infectent d'autres tissus.

La phosphoprotéomique peut contribuer à la découverte de nouveaux mécanismes responsables de la prolifération anarchique cellulaire menant à la formation de tumeurs. Les approches phosphoprotéomiques quantitatives permettent l'investigation de voies de signalisation activées de manière anormale pouvant servir de cibles dans le traitement du cancer. Ces approches permettent non seulement de déterminer des cibles thérapeutiques mais identifient aussi les effets « hors-cible » et estiment la toxicité suite à l'utilisation d'inhibiteurs par exemple. Des études phosphoprotéomiques visant la dissection moléculaire de plusieurs voies impliquées dans le cancer ont été publiées comme dans le cas de la voie EPHB2 [Ephrin type-B receptor 2][16], ATM [Ataxia telangiectasia-mutated]/ATR [ATM and Rad3-related][17], HER2/NEU[18] ou EGFR [récepteur du

facteur de croissance épidermal][19]. La majorité des expériences est réalisée dans des modèles cellulaires en culture mais des tissus tumoraux tels que des mélanomes peuvent être étudiés[20]. L'activation aberrante de kinases est souvent associée à divers cancers dans des organes tels que le poumon[21], le sein[22], le pancréas[23], la prostate[24] ou encore des tumeurs gastro-intestinales stromales[25] ou hématologiques[26]. L'utilisation d'inhibiteurs ciblant les kinases est une stratégie intéressante lors d'une approche pharmacologique[27]. Diverses molécules sont utilisées à des fins cliniques et de nombreuses autres sont encore en phase d'essai. Différents criblages ont été effectués afin de tester des composés chimiques en vue de l'élaboration de traitements thérapeutiques. C'est le cas notamment d'une étude menée afin de déterminer les effets de U0126 et SB202190 qui sont des inhibiteurs de MAPK[28].

Les voies associées aux récepteurs tyrosine kinase sont très souvent impliquées dans le cancer et leur dissection permet de mieux comprendre la complexité des mécanismes impliqués en aval[29]. Des inhibiteurs utilisés en clinique ciblent directement certains RTKs tels le gefitinib[30] ou encore l'erlotinib[31] (inhibiteurs de l'EGFR) utilisés dans des cas de cancer du poumon. Ces molécules rentrent en compétition avec l'ATP en se liant sur le récepteur. De même, le pazopanib est un inhibiteur du VEGFR (récepteur du facteur de croissance épidermal vasculaire)[32] et est utilisé pour le traitement du cancer du sein.

L'étude de la voie de l'insuline fait intervenir de nombreux événements de signalisation suite à la liaison de l'insuline sur son récepteur InR appartenant à la famille des RTKs. On considère que le RTK anormalement activé se comporte comme une oncoprotéine. Dans certains types de cancers, il est suggéré que plusieurs RTKs sont activés simultanément ce qui augmente la diversité des effecteurs et induit une transduction de signal plus efficace[33].

L'étude de la voie de signalisation de l'insuline et du rôle des RTKs permettent également de mieux comprendre les mécanismes impliqués dans des maladies métaboliques. En effet, la cascade de phosphorylation induite dans cette voie peut nous donner des indices sur le phénomène de résistance à l'insuline qui provoque une

dérégulation du taux de glucose dans le sang (diabète de type 2). Un défaut de production d'insuline conduit à un diabète de type 1. En 2010, on évaluait le nombre de personnes souffrant du diabète au Canada et dans le monde à respectivement 2 et 150 millions (source : Santé Canada). Environ 90 % des personnes concernées ont un diabète de type 2. Les problèmes d'obésité encourus suite à la dérégulation de cette voie sont souvent caractéristiques de ce syndrome métabolique. La présence de taux élevés d'acides gras libres produits par les adipocytes influence négativement l'absorption du glucose, la synthèse du glycogène ainsi que la glycolyse. La compréhension des processus moléculaires mis en jeu suite à l'activation de l'InR peut nous donner des indices cruciaux afin de cibler les interacteurs clés et développer un traitement efficace contre le diabète.

1.4. La drosophile comme modèle d'étude

Drosophila melanogaster, plus communément connue sous le nom de mouche du vinaigre ou mouche à fruit, est un des organismes modèles les plus utilisés en recherche biologique. Bien que le terme général « drosophile » puisse désigner plusieurs espèces, je l'utiliserais dans cette thèse pour décrire spécifiquement *Drosophila melanogaster*.

D'un point de vue de l'homologie génétique avec l'humain, la drosophile contient 61 % des gènes associés à des maladies connues et plus de la moitié des protéines exprimées chez la mouche ont des homologues chez l'humain. De plus, la voie de signalisation de l'insuline est très conservée par rapport aux mammifères. Il faut noter que le génome de la drosophile comporte moins de redondance fonctionnelle en comparaison avec le mammifère ce qui permet d'inhiber la fonction d'un seul gène seulement pour des études de perte de fonction.

1.4.1. Les cellules S2 de drosophile

Les cellules Schneider 2 (S2) de *Drosophila melanogaster* proviennent d'une lignée cellulaire dérivée d'une culture primaire d'embryons à stade tardif (âgés de 20 à 24 heures)[34]. De nombreuses caractéristiques telles que leur capacité de phagocytose suggèrent qu'elles sont dérivées d'une lignée de type macrophage. Les S2 sont un modèle cellulaire très populaire utilisé en biologie ou en génomique fonctionnelle. En effet, elles sont particulièrement faciles à cultiver et leur temps de division est plutôt rapide (environ 15 heures). Les S2 sont propices à l'utilisation de RNAi en vue d'inhiber l'expression d'une protéine. En effet, aucun agent de transfection n'est requis afin d'introduire l'ARN double brin dans les cellules ce qui permet un haut taux de transfection et donc maximise les chances d'atteindre une inhibition optimale en l'espace de quelques jours[35]. L'avantage réside également dans le fait que contrairement aux cellules de mammifères, la préparation de dsRNA peut se faire au laboratoire, ce qui réduit les coûts associés à l'achat et à la synthèse de siRNA.

1.4.2. La stratégie RNAi dans les cellules S2

Une des meilleures stratégies utilisées dans le but de déterminer la fonction d'une protéine est de l'éliminer au sein du système biologique d'intérêt afin d'y associer un phénotype. Plusieurs moyens sont envisageables dont notamment des approches de mutations génétiques assez laborieuses. Dans la dernière décennie, le développement de méthodes d'inhibition de protéines par ARN double brin s'est révélé un outil puissant permettant la caractérisation de protéines au niveau cellulaire. Cette technologie innovante a d'ailleurs fait l'objet du prix Nobel de Physiologie ou Médecine en 2006, récompensant les travaux de Andrew Z. Fire et Craig C. Mello chez *C. Elegans*[36]. Brièvement, l'ajout d'ARN double brin composé d'une séquence spécifique bloque la production d'une protéine cible et induit une perte de fonction partielle causée par le renouvellement des constituants protéiques au cours des divisions. L'ARN exogène introduit fait intervenir les

mêmes mécanismes immunitaires mis en jeu que lors d'une attaque virale, étant donné que le seul ARN présent dans la cellule devrait être simple brin. Chez la drosophile, l'ARN double brin est détecté et se lie à une protéine effectrice R2d2 qui stimule l'activité de la protéine ribonucléase Dicer-2. Dicer-2 clive de longues séquences d'ARN double brin sous forme de fragments d'une longueur de 21 à 25 paires de base (siRNA)[37]. L'hétérotrimère composé de siRNA, Dicer-2 et R2d2 est requis pour l'assemblage du complexe RISC (RNA-Induced Silencing Complex) ainsi que pour son activation. Après déroulement, les fragments sont séparés sous forme simple brin et intégrés au complexe RISC. Seul un des deux brins sert de guide et se lie à la protéine Ago-2 tandis que l'autre est dégradé. Ago-2 est une endonucléase du complexe RISC qui est nécessaire au clivage de l'ARNm[38]. Le complexe RISC activé et lié au siRNA s'apparie de manière spécifique à l'ARNm correspondant puis le clive, empêchant alors sa traduction en une protéine fonctionnelle[39]. Un schéma de ce processus est présenté en **Figure 1.4**.

Le protocole traditionnel utilisé pour inhiber l'expression d'une protéine dans des cellules S2 en phase exponentielle de croissance consiste à ajouter directement de courtes séquences d'ARN double brin (21 paires de bases en moyenne) dans le milieu de culture. Ces séquences vont entrer par l'intermédiaire de la voie de l'endocytose et font intervenir des protéines permettant la formation de vésicules telles que la clathrine ou encore des récepteurs de type éboueurs (scavengers)[40]. Cependant le mécanisme exact reste mal compris.

Bien que l'effet du RNAi ne soit pas considéré comme une inhibition totale, il s'agit néanmoins d'un outil commode afin d'associer un rôle physiologique à une protéine. Des efforts considérables ont été réalisés afin de minimiser les effets hors-cibles en prenant soin de concevoir des dsRNA avec les séquences les plus spécifiques et appropriées afin d'éviter l'inhibition d'autres protéines simultanément[41]. De plus, la viabilité cellulaire doit être préservée afin d'obtenir des réponses physiologiques fiables.

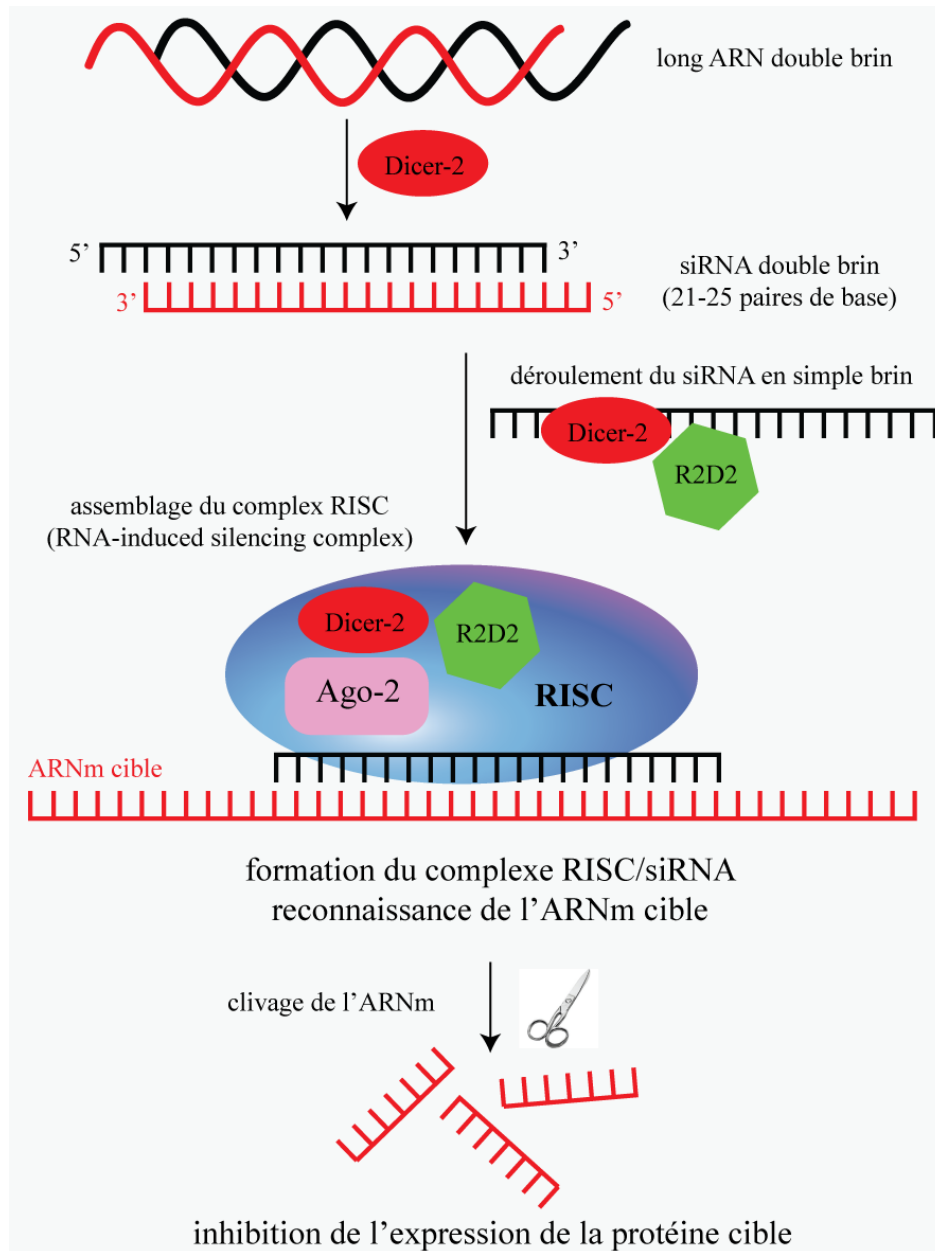


Figure 1.4. Représentation du mécanisme d'action de l'inhibition par ARN interférent chez la drosophile.

L'enzyme Dicer-2 clive l'ARN double brin exogène sous forme de courts fragments de 21 à 25 paires de base (siRNA). Les siRNA sont incorporés au sein du complexe RISC et s'apparient à la séquence complémentaire d'ARNm, induisant son clivage et l'inhibition subséquente de la traduction de la protéine cible.

Les cellules traitées avec le RNAi visant l'inhibition d'une protéine d'intérêt doivent être comparées avec des cellules traitées avec une séquence RNAi contrôle n'induisant aucune inhibition (séquence synthétisée au hasard ou de manière inverse, séquence visant la GFP exprimée uniquement chez la méduse...). Ceci permet de soustraire l'effet non spécifique lié à l'introduction d'une grande quantité de dsRNA dans les cellules. Souvent, différentes séquences de RNAi sont introduites simultanément afin de maximiser l'inhibition. L'utilisation d'inhibiteurs chimique est également envisageable pour empêcher l'expression d'une protéine mais il faut s'assurer de la non-toxicité du composé, de sa spécificité et surtout de sa disponibilité commerciale. Les effets hors-cibles sont envisageables aussi bien lors de l'utilisation de RNAi que lors du traitement avec de petites molécules inhibitrices. Théoriquement, n'importe quelle protéine peut être inhibée par RNAi une fois les séquences spécifiques déterminées et synthétisées. Cependant, il est difficile d'obtenir une très bonne inhibition pour des protéines très abondantes.

1.5. La voie de signalisation de l'insuline

1.5.1. Le rôle de l'insuline dans le métabolisme

Chez le mammifère, l'insuline est une hormone qui est produite dans le pancréas (par les cellules β des îlots de Langerhans plus précisément) afin de réguler le métabolisme des sucres et des lipides. Elle est stockée dans des vacuoles et a pour fonction première de stimuler le transport des nutriments (glucose, acides aminés et acides gras) du sang vers les muscles, les tissus adipeux et le foie. Son rôle majeur est d'éviter un taux excessif de glucose toxique dans le sang (hyperglycémie) mais elle régule également la conversion des nutriments en molécules de stockage telles que le glycogène, les protéines et les lipides. Une incapacité à synthétiser l'insuline résulte en un diabète de type 1. Une résistance à l'insuline ou une dysfonction des cellules β cause le diabète de type 2 et est induit suite à

un défaut de signalisation dans la voie. Lorsque le niveau de glucose est trop bas, le corps utilise les sucres stockés comme source d'énergie grâce à la glycogénolyse qui utilise le glycogène du foie et des muscles. En plus de favoriser le stockage du glucose, l'insuline inhibe sa production et sa libération par le foie en bloquant la gluconéogenèse et la glycogénolyse[42]. L'insuline contrôle directement l'activité de nombreuses enzymes métaboliques en modulant leur phosphorylation et en régulant l'expression de gènes codant pour des enzymes hépatiques.

La voie de signalisation de l'insuline est bien conservée entre les espèces. En effet, la drosophile a également besoin de mécanismes pour réguler son niveau de sucre ou stocker l'excès d'énergie sous forme de glycogène[43] et la plupart de ces processus fait intervenir la voie de l'insuline. Bien que les insectes ne possèdent pas de tissus tels que le pancréas spécialisés dans la régulation des carbohydrates, la drosophile possède un groupe de cellules produisant de l'insuline (IPC) localisé dans le cerveau qui a une fonction endocrinienne[44]. La drosophile a ce qu'on appelle un « corps gras » qui remplace le foie ainsi que des tissus adipeux qui sont le siège de la régulation métabolique. L'hormone adipocinétique (AKH) présente une homologie structurale avec le glucagon chez les vertébrés et joue un rôle déterminant dans la libération des sucres et des lipides stockés dans le corps gras sous forme de tréhalose et de diacylglycérol qui sont les deux sources d'énergie utilisées chez la drosophile. La libération d'AKH par le corps cardiaque est modulée par le taux de sucres en circulation suite à un changement dans la concentration intracellulaire en calcium. Celle-ci est médiée par des canaux ATP-dépendants sensibles au potassium (de manière similaire au mécanisme qui contrôle la sécrétion de glucagon par les cellules pancréatiques β)[45]. Ces observations démontrent que les fonctions métaboliques sont conservées à travers l'évolution et que la drosophile est un modèle valide dans le but d'étudier leurs mécanismes. En effet, chez les drosophiles mutantes pour InR, on note une augmentation de l'accumulation de triglycérides, ce qui est en accord avec le rôle de l'insuline dans la régulation du stockage de l'énergie en réponse à la disponibilité des nutriments. Un excès de nutriments active cette voie et engendre l'expansion de la masse grasseuse de l'insecte en augmentant le nombre d'adipocytes (ce qui est contrôlé

notamment par la phosphorylation du facteur de transcription Foxo [forkhead box protein O]). Ce phénomène est causé aussi par l'accumulation de masse lipidique qui est régulée par shaggy (homologue de GSK-3)[46] et dont la fonction sera décrite plus en détail dans la **section 1.5.2.2.**

1.5.2. L'activation de InR et la cascade de phosphorylation en aval

Chez la drosophile, 7 peptides ressemblant à l'insuline nommés Ilp1-7 [Insulin-like protein][47] sont décrits dans la littérature et Ilp2 est celui qui se rapproche le plus de l'insuline chez le mammifère. La voie de l'insuline est impliquée dans de nombreux processus tels que la régulation du métabolisme, la reproduction et la détermination de la durée de vie[48, 49] mais également dans la régulation de la croissance et de la taille de l'organisme[50]. La protéine Insulin-like receptor (InR) est le RTK homologue du récepteur de l'insuline chez le mammifère. InR a une taille d'environ 400 kDa et comporte deux sous-unités α (120 kDa) et deux sous-unités β (95 kDa) incluant une tyrosine kinase cytoplasmique sur la sous-unité β qui est activée suite à la liaison de l'insuline[51] (**Figure 1.5**). La seule différence dans sa structure, comparativement aux mammifères, réside en un ajout de 398 acides aminés supplémentaires contenant des motifs YXXM semblables à ceux trouvés dans IRS-1 (Insulin Receptor Substrate 1) chez l'humain[52]. L'insuline se lie à la portion extracellulaire du récepteur sur les sous-unités β . Ceci induit un changement conformationnel qui active le domaine kinase sur la portion intracellulaire. Ce domaine kinase induit ensuite l'autophosphorylation des résidus tyrosines en C-ter du récepteur ce qui déclenche la cascade de phosphorylation. Le degré d'homologie de séquence pour ce récepteur est très grand surtout pour le domaine kinase.

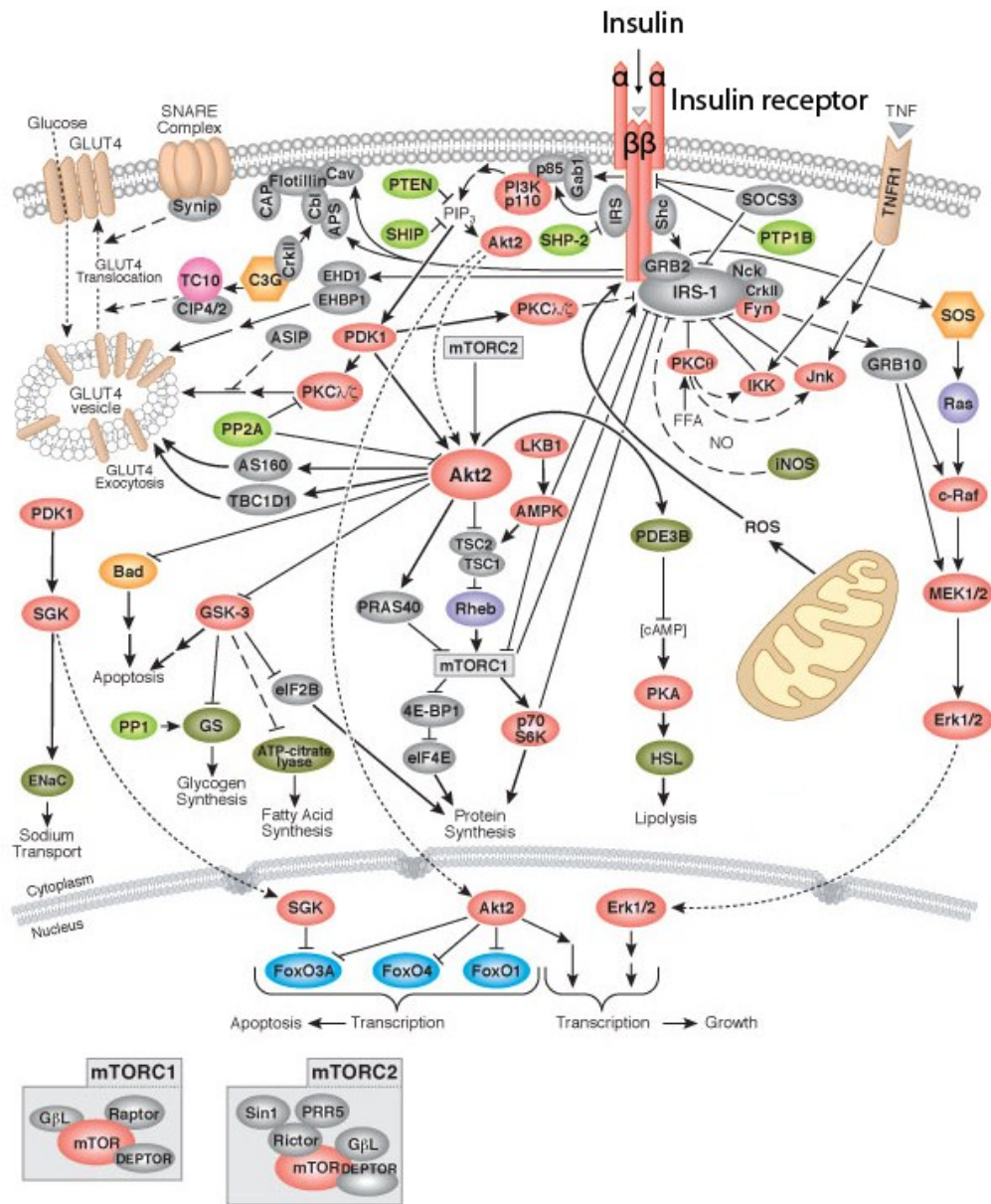


Figure 1.5. Schéma global des voies de signalisation activées par l'insuline.

L'insuline se lie au domaine extracellulaire du récepteur induisant un changement de conformation puis son autophosphorylation sur les tyrosines en C-ter sur le domaine intracellulaire. De nombreuses voies sont activées telles que la voie de signalisation métabolique impliquant PI3K et Akt, la régulation du glycogène via GSK3, la régulation du taux de glucose via GLUT4 ou encore la voie mitogénique faisant intervenir Ras/MAPK. Modifié à partir de : http://www.cellsignal.com/reference/pathway/Insulin_Receptor.html avec la permission de Cell Signaling Technology, copyright 2010.

La voie de l'insuline fait intervenir une cascade de phosphorylation hiérarchique (**Figure 1.5**). Chez la drosophile, il y a un homologue pour pratiquement tous les composants de la voie de l'insuline mais la complexité globale du système est réduite par rapport aux mammifères qui peuvent avoir plusieurs isoformes. Il est suggéré qu'InR est impliqué dans plusieurs voies de transduction de signal décrites dans les sections suivantes.

1.5.2.1. La voie Ras-MAPK

La voie Ras-MAPK implique la liaison directe du récepteur après son autophosphorylation avec des protéines contenant un domaine SH2 ou SH3 (Src homology 2 ou 3) ou PTB (liaison aux phosphotyrosines). La protéine adaptatrice Grb2 liée à IRS1 s'associe avec Sos (son of sevenless) afin d'activer la voie Ras/MAPK[53] suite à sa liaison avec la petite GTPase Ras. Ras subit alors un changement de conformation permettant le remplacement du GDP lié par du GTP puis son activation. SHC (Src Homology 2 domain Containing) est également un substrat de InR qui peut s'associer à Grb2 de manière indépendante à IRS. Ces événements engendrent la cascade de phosphorylation hiérarchique impliquant Ras, Raf, Mek (Map kinase/Erk Kinase) et Erk (Extracellular signal-regulated kinase). Suite à sa phosphorylation, Erk est transloqué dans le noyau afin de médier la réponse mitogénique via la transcription de gènes. Le rôle direct de rolled (homologue d'Erk) dans la prolifération cellulaire et l'apoptose des cellules S2 traitées à l'insuline a déjà été mis en évidence[54, 55]. La voie de l'insuline a été étudiée en détails lors d'une étude par RNAi et a démontré l'importance de rolled dans la transduction de signal[56]. Elle est connue pour être phosphorylée sur les sites T₂₄₈/Y₂₅₀ (boucle d'activation) par Mek en réponse à l'insuline. Par ailleurs, des phénomènes de rétroaction ont été observés dans le cas de la voie MAPK qui agit via le facteur de transcription pointed afin de contrôler l'expression du gène Inr[57].

Globalement, les MAPKs sont des kinases qui transforment des stimuli extracellulaires en réponses intracellulaires par le biais de récepteurs induisant une

transduction de signal hiérarchique. Les récepteurs impliqués peuvent être couplés à des protéines G, à des cytokines, à des facteurs de croissance ou encore associés à des stress environnementaux. Les MAPKs phosphorylent des MAPK-APKs (protéines kinases activées par les MAP-kinases)[58] et sont phosphorylées par des MAPKKs (MAP Kinase Kinases telles que MEK). Elles peuvent être déphosphorylées et inactivées par plusieurs (MAP kinases)-phosphatases (MKPs). Elles font partie d'une cascade de phosphorylation impliquant différents interacteurs qui propagent le signal jusqu'à ce qu'il soit transmis adéquatement à travers la cellule. La famille des MAPKs comporte 4 catégories : MAPK P38, JNK (c-Jun N-terminal Kinase) (JNK1, JNK2 et JNK3), ERK et ERK5. Leur mode de fonctionnement est régulé par des interactions d'empilement (docking interaction) qui régissent les interactions protéine-protéine. Ces sites d'interaction sont situés en dehors du domaine catalytique et sont constitués de courts segments d'acides aminés (motif conservé) proches de l'extrémité N-ter présents à la fois sur les MAPKs et les substrats avec lesquelles elles interagissent[59]. Des protéines d'échafaudage (scaffold proteins) sont également nécessaires au fonctionnement des MAPKs telles que JIP (JNK-Interacting Protein) et MKK7 (Map Kinase Kinase 7).

1.5.2.2. La voie de PI3K et AKT

La seconde voie permettant la transduction du signal de l'insuline implique la sous-unité β tronquée et induit la phosphorylation de la protéine adaptatrice Chico (IRS-1) ou Lnk (SH2b) pour transmettre le signal[60, 61]. Bien que de nombreuses protéines adaptatrices soient décrites dans la littérature comme liant le récepteur chez les mammifères (IRS-1, IRS-2, IRS-3, IRS-4, SHC, CBL [Casitas B-lineage lymphoma proto-oncogene], APS [Adapter protein with pleckstrin homology and Src homology 2 domains], SH2B, GAB1 [Growth factor receptor bound protein 2-associated protein 1], GAB2, DOCK1 [Dedicator of cytokinesis protein 1], DOCK2 et CEACAM-1 [Carcinoembryonic antigen-related cell adhesion molecule])[62], seules les protéines Chico et Lnk ont été mises en évidence chez la drosophile. Chico se lie à Pi3k92e qui catalyse la formation de PI(3,4,5)P3

(phosphoinositide 3,4,5 triphosphate) puis active Akt et Pdk1 [Phosphoinositide-dependent kinase 1][63]. Leur activité est aussi régulée via la phosphatase de lipides Pten (Phosphatase and tensin homolog) qui est un suppresseur de tumeur agissant négativement dans la voie[64]. Il est connu que l'activité d'Akt est modulée par sa déphosphorylation par Pp2a. Akt est phosphorylée sur la S₄₇₃ en réponse à l'insuline par Tor et sur la T₃₀₈ par Pdk1[65]. Elle module les activités enzymatiques qui affectent la génération d'oxyde nitrique et l'apoptose, contrôle le niveau de glucose et régule le métabolisme des lipides et des protéines. Tor est une protéine reliée à la famille des kinases de type phosphatidylinositol et est aussi impliquée dans le contrôle de la taille en réponse au taux de nutriments[66]. Elle a pour cible S6k qui phosphoryle ensuite la protéine ribosomale S6 impliquée dans la régulation de l'ARNm. Les gènes suppresseurs de tumeurs Tsc1 (Tuberous Sclerosis Complex 1) et Tsc2 (Tuberous Sclerosis Complex 2) interviennent aussi dans la voie de signalisation de l'insuline et ont un effet global antagoniste[66]. Il est connu qu'Akt peut déstabiliser le complexe Tsc1/Tsc2 étant donné que la phosphorylation de Tsc2 par Akt le rend inactif (en condition de culture cellulaire). La connexion entre Tsc1/Tsc2 et Tor a été reliée génétiquement à Rheb (Ras-homologue expressed in brain) qui est une petite protéine liée au GTP. Tsc2 agit comme une GAP (GTPase-activating protein) et inhibe Tor-C1 via Rheb. Inversement, l'inhibition de Tsc2 induit l'activation de Tor et module la croissance des tissus[67]. Le facteur de transcription Foxo est décrit comme une autre cible de Akt suite à l'insuline ce qui force sa rétention dans le cytoplasme et inhibe son activité transcriptionnelle[68]. De cette manière, il a tendance à induire une conservation d'énergie et un arrêt de la croissance. Le répresseur de traduction 4E-BP1 (EIF4E-binding protein 1) (Thor) est une cible de Foxo qui après activation se lie à Eif4E (Eukaryotic initiation factor 4E) et bloque la traduction[69].

L'insuline a un effet sur la synthèse protéique en stimulant l'entrée d'acides aminés dans la cellule, en inhibant la dégradation de protéines et favorisant leur production. Suite à son activation, AKT est transloqué dans le cytoplasme puis phosphoryle et inactive GSK3 (Glycogen Synthase Kinase 3) (S₂₁ pour GSK3 α et S₉ pour GSK3 β)[70]. Chez la souris, il a été prouvé que GSK3 β a un rôle prédominant[71]. GSK3 est active de manière

constitutionnelle et un de ses substrats majeurs est l'enzyme glycogène synthase dont la phosphorylation empêche la synthèse de glycogène. Ainsi, l'inactivation de GSK3 par AKT conduit à l'accumulation et au stockage du glucose sous forme de glycogène. L'insuline inhibe la capacité de GSK3 à phosphoryler ses substrats (qui sont de préférence déjà phosphorylés[72]) en induisant sa phosphorylation et donc en bloquant les interactions potentielles à cause d'un encombrement stérique[73]. GSK3 inactivée engendre la déphosphorylation de facteurs de transcription tels que CBP, c-Myc, c-Jun[74] et de facteurs d'initiation de la traduction tels que Eif2B (Eukaryotic initiation factor 2B) ce qui favorise la synthèse de protéines et le stockage d'acides aminés[75]. L'inactivation de GSK3 peut également avoir des effets antiapoptotique[76]. Par ailleurs, GSK3 phosphoryle également IRS1 (S₃₃₂) et IRS2 (S₄₈₄), engendrant l'inhibition de la voie de l'insuline par rétroaction négative impliquant également PKC β et JNK[77].

1.5.2.3. La régulation du taux de glucose via GLUT4

Un des modes d'action de l'insuline pour stimuler la consommation de glucose par les cellules est d'induire la translocation d'un transporteur GLUT4 (Glucose transporter type 4) du milieu intracellulaire vers la membrane plasmique par l'intermédiaire de PI3K et AKT. La stimulation à l'insuline a pour effet d'augmenter l'exocytose des vésicules contenant GLUT4 et diminue légèrement le taux d'internalisation[78]. Un mécanisme indépendant met également en jeu la phosphorylation du proto-oncogène CBL qui s'associe à la protéine CAP (adenylyl cyclase-associated protein 1) puis le complexe est transloqué vers des radeaux lipidiques à la membrane. CBL interagit ensuite avec la protéine adaptatrice CRK associée à C3G appartenant à la famille des Rho-GEFs (Rho-family guanine nucleotide exchange factor). C3G active ensuite TC10 (protéine se liant au GTP) ce qui engendre la translocation de GLUT4 à la membrane selon un mécanisme encore mal compris[75]. Ces événements seraient dépendants du cytosquelette et feraient intervenir un réseau de microtubules impliquant des kinésines, des filaments d'actine ou encore une interaction avec VAMP2 (Vesicle-associated membrane protein 2) et MUNC18c

(Mammalian uncoordinated-18c) pour la formation de complexes SNARE (SNAP [Soluble NSF attachment protein] receptor)[79].

Il est cependant difficile de mesurer les effets directs de l'insuline chez la drosophile en condition d'excès de nutriments pour l'étude du diabète. Il est tout de même envisageable de faire des analogies de mécanismes vis-à-vis, par exemple, de la résistance centrale à l'insuline induite par des taux élevés de triglycérides chez le mammifère. Ceci est ensuite interprété par l'hypothalamus comme une faible disponibilité en nutriments et permet d'ajuster les processus physiologiques pour s'adapter au déficit et réduire la division cellulaire.

1.5.3. Les études phosphoprotéomiques de la voie de l'insuline

Un des objectifs de cette thèse est d'explorer de manière globale les événements de phosphorylation régulant la voie de l'insuline chez la drosophile (**Chapitre 2**). Des études ont déjà été rapportées dans la littérature afin de disséquer cette voie de signalisation par des approches phosphoprotéomiques plus ou moins exhaustives. Une étude publiée par Schmelzle, *et al.* en 2004 a répertorié la cinétique de phosphotyrosines régulées par l'insuline dans des adipocytes blancs 3T3-L1 grâce à une approche combinant immunoprécipitation et enrichissement de phosphopeptides sur IMAC (chromatographie d'affinité au métal immobilisé)[80]. La quantification a été réalisée par i-TRAQ (étiquette isobarique pour la quantification relative et absolue de peptide). Une autre étude en 2008 a été conduite en combinant une étape d'immunoprécipitation avec une quantification par SILAC (marquage isotopique stable en culture cellulaire) lors d'une cinétique réalisée dans des cellules d'adipocytes bruns différenciées chez la souris. Une quarantaine de sites de phosphorylation sur des tyrosines ont été décrits comme modulés suite à une stimulation à l'insuline[81]. Une étude plus ciblée a étudié les protéines modulées par l'insuline et régulées par AKT en se basant sur les protéines se liant à la protéine kinase 14-3-3 par une approche utilisant SILAC[82]. Une analyse des complexes formés autour de la protéine

Chico après immunoprécipitation dans des cellules de drosophile stimulées ou non à l'insuline a été conduite et a également mis en évidence plusieurs sites de phosphorylation sur Chico et InR[83]. *In vivo*, une analyse phosphoprotéomique à grande échelle a été réalisée en utilisant SILAC combiné à l'ajout de phosphopeptides isotopiquement marqués afin d'analyser les changements de phosphorylation dans des foies de souris[84].

Dans cette thèse, nous proposons une étude exhaustive quantitative et exploratoire de la voie de l'insuline chez la drosophile dans des cellules S2 en culture ce qui n'a jamais été décrit auparavant.

1.6. Protéine kinase CK2

La protéine kinase CK2 (ou caséine kinase 2) est une kinase spécifique aux sérines et aux thréonines exprimée à la fois dans le cytoplasme, dans le noyau ainsi que dans divers organites[85, 86]. Elle a aussi la capacité de phosphoryler des résidus tyrosines dans certains cas[87, 88] et utilise soit l'ATP, soit le GTP comme donneurs de phosphate. Chez le mammifère, elle est constituée d'un hétérotétramère formé de deux sous-unités catalytiques (α et α' de 42 et 38 kDa respectivement) et de deux sous-unités β (28 kDa) requises pour la reconnaissance de son substrat, son activité et sa stabilité[89-92]. CK2 phosphoryle ses substrats protéiques sur des sites à proximité de résidus acides dans un motif consensus général S*/T* X X D/E[93].

CK2 est impliquée dans la croissance et la prolifération[91], la progression du cycle de division et la viabilité cellulaire[94, 95]. Elle a une fonction majeure dans la tumorigenèse[96] et dans le contrôle de la mort cellulaire en ayant un rôle de suppresseur de l'apoptose[97, 98]. Par exemple, l'activation de NF- κ B est déclenchée par la phosphorylation de I- κ B par diverses kinases telles que CK2, ce qui lève l'inhibition du complexe et mène à sa dégradation. Ensuite, NF- κ B est transloqué au noyau puis se lie à des sites de régulation pour moduler l'activité de plusieurs gènes[99]. Par ailleurs, la phosphorylation de certaines protéines telles que MAX (Myc-associated factor X) par CK2

permet de la protéger de la dégradation via le clivage médié par les caspases[100]. CK2 est connue pour phosphoryler des centaines de substrats[101]. De nombreuses cibles de CK2 sont rapportées chez le mammifère comme les proto-oncogènes c-MYC[102], c-MYB[103] ou encore c-JUN[104], des gènes suppresseurs de tumeurs comme p53[105] et BRCA1 (Breast cancer type 1 susceptibility protein)[106] et des régulateurs de la transcription tels que MAX[107], CUT[108], IRF4 (Interferon Regulatory Factor 4)[109], SIX1[110] et TOP2 (Topoisomerase 2)[111]. Chez la drosophile, CK2 a un rôle déterminant dans le développement et la différenciation en modulant Deadpan[112], Hairy[113] ou encore Cactus[114] ainsi que dans la régulation de gènes de transcription en phosphorylant GAGA[115], Groucho[116], Fmr1[117], Nap1[118] et Hp1[119].

D'une manière générale, les niveaux observés de CK2 sont proportionnels au taux de prolifération cellulaire[120] ce qui explique pourquoi des taux élevés de CK2 ont été mesurés dans divers cas de cancers[121]. Bien que CK2 ait été découverte pour la première fois il y a 60 ans, les mécanismes qui régissent son activité restent encore mystérieux. On pense que CK2 est active de manière constitutionnelle[122, 123] et ne requiert pas de phosphorylation pour être fonctionnelle contrairement à d'autres kinases, mais ce point de vue est encore discuté. La littérature s'entend pour dire plutôt qu'une certaine population de CK2 serait active en tout temps tandis qu'une autre partie serait activée localement par divers stimuli[124]. L'assemblage des sous-unités ainsi que les interactions avec diverses molécules et protéines cellulaires contribuent à sa régulation. Il est suggéré que la localisation intracellulaire de CK2 influence son activité[86]. Certaines études ont démontré une augmentation de son activité suite à une stimulation au sérum[125], à l'insuline ou à un traitement à l'IGF[126, 127] ou encore à l'EGF[128], mais les conclusions sont encore débattues[129]. Il a été démontré que CK2 est impliquée dans la régulation de la voie de l'insuline dans des cellules pancréatiques de type β [130]. Une augmentation de la phosphorylation de CK2 sur sa sous-unité β suite à un traitement à l'EGF dans des cellules de carcinome a aussi été rapportée[128].

Chez la drosophile, un isoforme de CK2 α (DmCK2a) et trois de CK2 β (DmCK2b, DmCK2b0 et DmCK2btes) ont été décrits dans la littérature ainsi que leur distribution spécifique dans différents tissus[131-133]. L'autophosphorylation de CK2 a déjà été observée où DmCK2b est phosphorylé par DmCK2a sur les sérines S₂ et S₄ *in vitro*[134]. Une analyse fonctionnelle a été conduite sur DmCK2b afin de déterminer son rôle plus en détail[135]. Chez l'humain, la sous-unité β de CK2 est également phosphorylée différemment au cours du cycle cellulaire par CDC2 (Cell division control protein 2)[136]. La phosphorylation de la sous-unité α est aussi modulée selon le cycle de division sur sa partie C-ter.

CK2 est inhibée de manière générale par des composés négativement chargés tels que l'héparine et activée par des molécules positivement chargées comme les polyamines[89]. Divers inhibiteurs chimiques ont été développés pour CK2 et sont souvent des compétiteurs de l'ATP[137-139]. Le premier inhibiteur développé au stade clinique a été décrit par Siddiqui-Jain, *et al.* en vue d'un traitement contre le cancer[140]. L'utilisation de petites molécules ou de peptides bloquant la phosphorylation de CK2 a également été suggérée[141]. L'identification de substrats de CK2 peut permettre une meilleure compréhension de son mécanisme d'action afin de développer de nouvelles stratégies thérapeutiques.

1.7. La préparation d'échantillon en phosphoprotéomique

La réalisation d'une étude phosphoprotéomique à grande échelle requiert plusieurs étapes essentielles qui doivent être optimisées en fonction de l'échantillon à analyser. Un schéma global du protocole standard qui a été utilisé dans cette thèse est présenté en **Figure 1.6**.

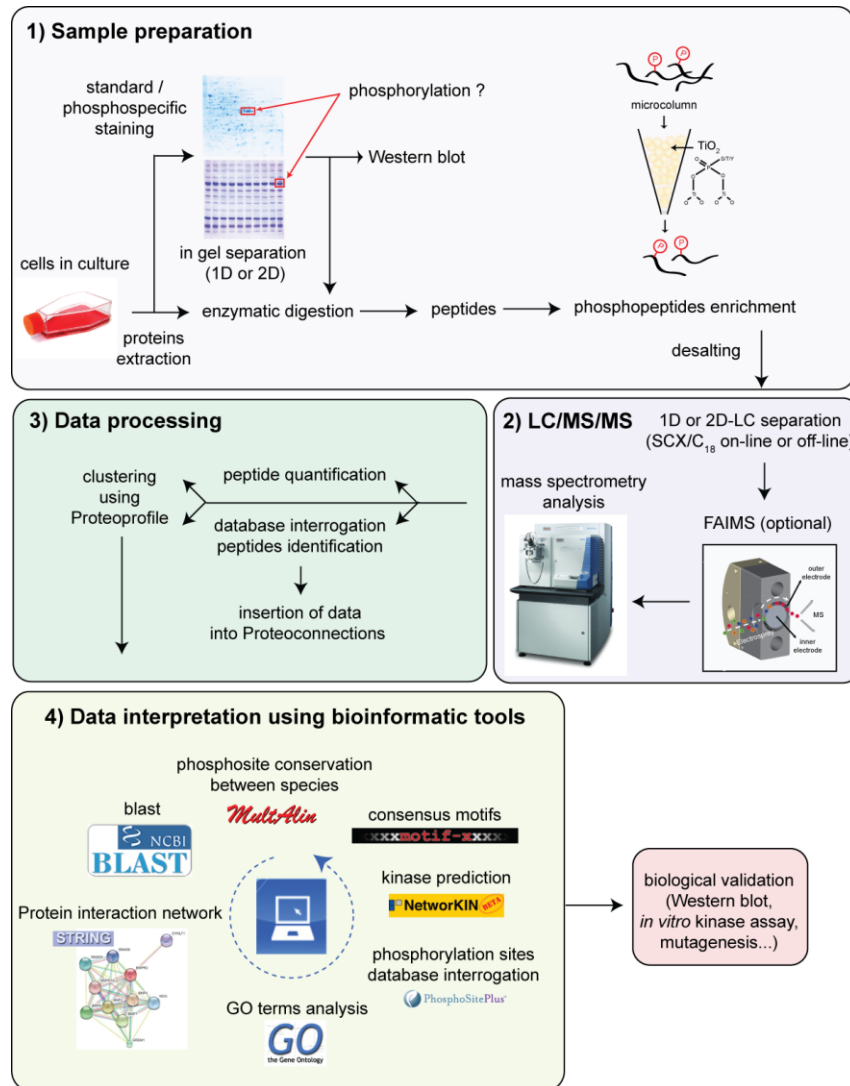


Figure 1.6. Schéma global du protocole opératoire utilisé dans cette thèse lors d'analyses phosphoprotéomiques.

1) La préparation de l'échantillon regroupe la culture cellulaire, l'extraction protéique, la digestion enzymatique et l'enrichissement des peptides en phosphopeptides. 2) Les phosphopeptides purifiés sont séparés par chromatographie liquide (puis de manière optionnelle par FAIMS) et analysés par spectrométrie de masse. 3) Les données enregistrées sont traitées afin d'obtenir les identifications protéiques suite à l'interrogation des bases de données et la quantification est réalisée grâce au logiciel Proteoprofile. 4) Les données sont interprétées grâce à l'utilisation de logiciels bioinformatiques afin d'émettre des hypothèses quant à la fonction biologique des sites de phosphorylation observés. Ces hypothèses sont ensuite validées par diverses approches biologiques et biochimiques.

1.7.1. La lyse cellulaire

Lors de l'étude d'un phosphoprotéome, la première étape consiste en l'isolation des protéines grâce à une lyse cellulaire appropriée (tampon salin hypotonique, bris mécanique dans un homogénéisateur, sonication...). Un fractionnement cellulaire est également possible lors d'études ciblant des compartiments plus précis[142]. Lors de la lyse, de nombreuses protéases et phosphatases actives dans la cellule sont susceptibles de dégrader l'échantillon. De plus, la stabilité de la phosphorylation est fragile et de nombreuses précautions doivent être prises afin de protéger son intégrité. L'ajout de mélanges d'inhibiteurs complets permettant l'inhibition d'une grande variété de réactions enzymatiques[143], la présence d'EDTA (acide éthylène diamine tétraacétique) et d'EGTA (acide éthylène glycol tétraacétique) ainsi que le travail à 4°C sont des conditions primordiales[144]. Par ailleurs, il est important de prendre soin d'utiliser des réactifs compatibles avec l'étude subséquente par spectrométrie de masse afin de minimiser les interférences en aval.

1.7.2. Digestion enzymatique

Après l'extraction de protéines suite à la lyse cellulaire, l'étape suivante consiste à effectuer une digestion enzymatique dans le but de générer des peptides analysables par spectrométrie de masse. Il est tout à fait possible d'analyser une protéine intacte puis de déterminer sa masse exacte, cependant plusieurs raisons favorisent l'analyse au niveau peptidique. En effet, les protéines ne sont pas toujours solubles dans certaines conditions en absence de détergent (qui est le plus souvent incompatible avec les analyses MS à cause de sa haute capacité d'ionisation qui interfère avec l'échantillon). De plus, afin d'obtenir de l'information sur la séquence peptidique et les modifications potentielles, il est indispensable d'obtenir une bonne fragmentation par MS/MS, celle-ci étant optimale pour des peptides de longueur inférieure à 25 acides aminés en général.

Chacune des enzymes utilisées en protéomique possède une spécificité particulière et clive la séquence au niveau d'acides aminés précis. La trypsine est l'enzyme la plus couramment utilisée car c'est une protéase stable avec une bonne activité qui clive les protéines après les résidus lysines et arginines (sauf s'ils sont suivis d'une proline qui crée un encombrement stérique)[145]. Cela engendre la formation de peptides d'une longueur acceptable (5-25 a.a.) afin d'obtenir un bon séquençage par MS/MS et qui comportent une charge positive grâce à la présence d'un résidu basique en C-ter. C'est pour ces raisons que c'est l'enzyme qui a été utilisée dans cette thèse. L'endoprotéase Lys-C clive la séquence peptidique après les résidus lysines et est encore plus stable que la trypsine. Elle génère des fragments plus longs et peut être utilisée seule ou en combinaison avec la trypsine du fait qu'elle garde son activité même dans des conditions plus drastiques. Les enzymes Asp-N et Glu-C sont elles aussi très spécifiques (mais moins actives) et clivent les séquences après les résidus acide aspartique et acide glutamique respectivement. Elles peuvent être utilisées afin de générer des peptides complémentaires à ceux obtenus avec la trypsine dans le but d'obtenir une meilleure couverture de séquence. Les autres protéases moins spécifiques sont à éviter car elles génèrent des fragments qui se chevauchent et rendent l'échantillon encore plus complexe. L'utilisation combinée de plusieurs protéases est idéale afin d'avoir une couverture maximale du protéome[146]. Cependant, cette stratégie n'a pas été appliquée aux projets présentés dans cette thèse à cause de la grande quantité de matériel requise et du temps d'analyse nécessaire.

1.7.3. Détection de phosphoprotéines par électrophorèse sur gel

1.7.3.1. Gel SDS-PAGE en une dimension

Une des méthodes permettant la visualisation et l'identification subséquente de phosphoprotéines consiste à les faire migrer sur un gel SDS-PAGE en une dimension en conditions dénaturantes. Cette technique permet la séparation des protéines selon leurs

poids moléculaires. On peut alors séparer un mélange complexe ou encore isoler une protéine provenant d'une étape d'immunoprécipitation. Si une protéine est phosphorylée, sa masse sera supérieure à sa forme non-phosphorylée (proportionnellement au nombre de groupement phosphates ajoutés) et une bande supérieure supplémentaire pourra être visible (dépendamment de son abondance). La détection peut se faire sur un gel coloré au bleu de Coomassie si l'abondance de la phosphorylation est suffisante (> 50 ng) ou encore au bleu de Coomassie colloïdal G-250 pour une détection plus sensible de l'ordre du ng[147]. La coloration au nitrate d'argent a une limite de détection comparable mais est déconseillée pour l'identification de sites de phosphorylation car elle peut causer des artéfacts confondus avec des événements de sulfatation provenant du protocole de coloration[148]. Des méthodes de coloration détectées par fluorescence telles que le Sypro-Ruby (Molecular Probes), Oriole et Flamingo (Bio-Rad), Lucy (Sigma-Aldrich), Deep Purple (GE Healthcare), Krypton (Thermo Scientific), Nimble Juice (GeneDirex) sont également des méthodes de choix pouvant détecter des bandes protéines de l'ordre du ng[149-151] mais nécessite l'acquisition d'un scanner compatible.

Une approche de migration sur un gel en deux dimensions est aussi possible afin d'identifier une protéine phosphorylée étant donné que la phosphorylation apporte une charge négative supplémentaire et modifie le point isoélectrique de la protéine.

1.7.3.2. **Immunobuvardage de type Western et approches utilisant des anticorps**

Une détection par immunobuvardage de type Western est envisageable grâce à l'utilisation d'anticorps reconnaissant de manière spécifique les phosphosérines, phosphothréonines et phosphotyrosines[152]. En pratique, seuls les anticorps anti-phosphotyrosines sont réellement fiables et produisent peu de bruit de fond. Plus précisément, il est envisageable de détecter directement la phosphorylation d'une protéine d'intérêt si un anticorps phosphospécifique est disponible. Malheureusement, leur accès est assez limité et leur spécificité varie selon la protéine d'intérêt et la compagnie qui le produit. La situation est

surtout problématique chez les organismes où les sites de phosphorylation ne sont pas toujours conservés par rapport aux mammifères (chez la drosophile par exemple). Il est alors possible de faire synthétiser un anticorps sur commande mais le résultat n'est malheureusement pas garanti et la procédure est longue et dispendieuse. Dans le cas où seulement un anticorps spécifique à la protéine existe, la détection d'une bande de plus haut poids moléculaire est possible si l'épitope reconnaissant la séquence spécifique de la protéine ne se situe pas dans la même région que le site de phosphorylation.

1.7.3.3. Coloration spécifiques aux phosphoprotéines

Des colorations plus ou moins sensibles destinées à marquer spécifiquement les phosphoprotéines sur gel ont été développées. La compagnie Pierce propose un kit de coloration spécifique pour détecter les phosphoprotéines qui colore en vert les bandes protéiques phosphorylées sur des sérines ou thréonines. La compagnie Invitrogen a développé la coloration Pro-Q[®] Diamond qui permet de détecter les phosphoprotéines modifiées sur les sérines, thréonines et tyrosines par fluorescence avec une sensibilité de 4 ng par spot. Cette technique a déjà été utilisée avec succès afin d'identifier des biomarqueurs dans des maladies neurodégénératives telles que la sclérose multiple[153] ou encore pour l'étude du phosphoprotéome de *Mycoplasma genitalium* et *Mycoplasma pneumoniae*[154]. La compagnie Sigma-Aldrich a développé le kit PhosDecor pour la détection de phosphoprotéines par fluorescence. D'une manière générale, ces techniques de coloration phosphospécifique sont des outils utiles pour détecter les événements de phosphorylation ayant une abondance raisonnable mais le bruit de fond reste tout de même un problème majeur.

Généralement, suite à la séparation de protéines par gel, la bande d'intérêt colorée est découpée puis la protéine est digérée de manière enzymatique et identifiée par spectrométrie de masse. Cette stratégie est également applicable lors de l'étude d'un protéome complet où la ligne de migration est découpée en plusieurs bandes de tailles

égales analysées séparément par MS. Cela permet de rendre l'échantillon moins complexe et d'identifier des protéines moins abondantes qui auraient été masquées par d'autres protéines majoritaires sans fractionnement.

1.7.3.4. **Autoradiographie**

Les phosphoprotéines peuvent être marquées de manière radioactive au ^{32}P ou ^{33}P [155, 156] et détectées par autoradiographie. C'est une méthode particulièrement sensible mais qui peut s'avérer non compatible avec la spectrométrie de masse. Il est possible d'incorporer un groupement orthophosphate ($\text{H}_3^{32}\text{PO}_4$) radioactif *in vivo* directement dans les cellules même si la présence d'ATP non marqué endogène peut interférer et résulter en une incorporation incomplète. Le processus peut être toxique pour les cellules et causer des dommages importants qui peuvent rendre le modèle expérimental non conforme. Il est possible *in vitro* d'effectuer des essais kinase avec du $[\text{g-}^{32}\text{P}]\text{-ATP}$ qui est incorporé et marque les résidus nouvellement phosphorylés par des kinases spécifiques ajoutées en excès.

1.7.4. **Les méthodes d'enrichissement en phosphoprotéines et phosphopeptides**

On estime que la proportion de protéines qui est sous forme phosphorylée à un moment donné est en moyenne de l'ordre de 0.1-1 %. En effet, bien que certaines protéines soient phosphorylées de manière constitutive, la plupart d'entre elles sont phosphorylées de façon transitoire, d'où leur faible abondance. Il est donc indispensable d'effectuer une étape d'enrichissement afin de pouvoir détecter des formes phosphorylées. Un schéma global des méthodes les plus courantes est présenté en **Figure 1.7**.

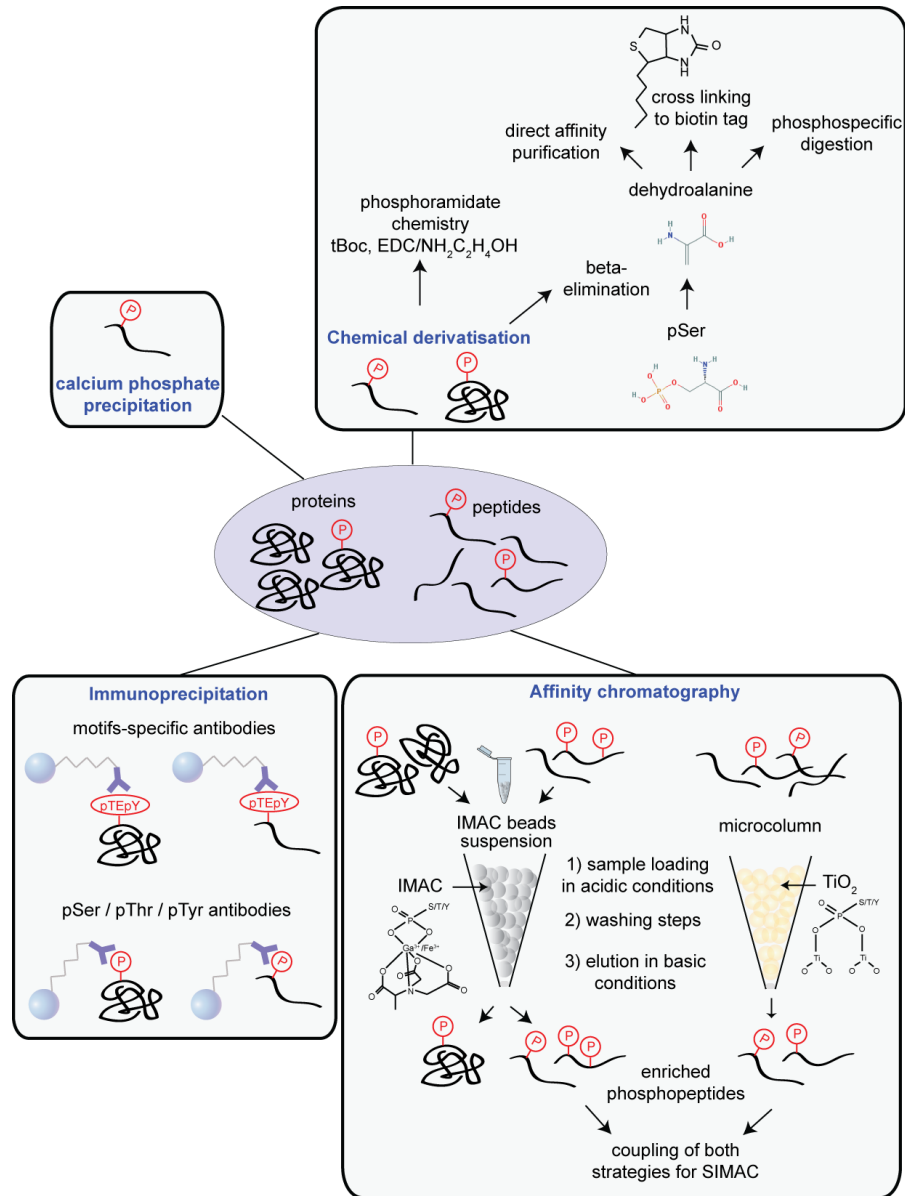


Figure 1.7. Schéma des stratégies les plus couramment utilisées pour l'enrichissement en phosphoprotéines et phosphopeptides.

Il est possible d'effectuer un enrichissement par immunoprécipitation à l'aide d'anticorps spécifiques aux pSer, pThr et pTyr ou à certains motifs reconnus par des kinases d'intérêt. La purification est aussi réalisable suite à une dérivatisation chimique ou une étape de précipitation au phosphate de calcium. L'approche la plus populaire reste la chromatographie d'affinité impliquant les phases IMAC et TiO_2 . D'autres types de phases stationnaires existent et sont décrites dans le texte.

1.7.4.1. Immunoprécipitation à l'aide d'anticorps

Il est possible d'enrichir un extrait protéique en phosphoprotéines grâce à une étape d'immunoprécipitation mettant en jeu des anticorps reconnaissant les phosphosérines et les phosphothréonines[157]. Leur spécificité n'est pas exceptionnelle mais permet quand même un certain enrichissement[158]. Les anticorps dirigés contre les phosphotyrosines sont plus prometteurs bien que leur phosphorylation soit moins fréquente[159]. Cette méthode fonctionne également pour l'enrichissement au niveau peptidique. En effet, Rush, *et al.* ont utilisé cette stratégie en utilisant des anticorps spécifiques aux phosphotyrosines dans des cellules cancéreuses 3T3-Src et ont mis en évidence 185 sites de phosphorylation[160]. De même, Villen, *et al.* ont identifié 385 phosphotyrosines phosphorylées lors de leur étude du phosphoprotéome du foie chez la souris[161].

Une autre approche utilisant des anticorps motif-spécifiques a déjà été publiée et consiste à purifier des phosphopeptides contenant une séquence consensus pour une kinase d'intérêt. Ce concept fut testé avec succès pour ATM ou ATR (anticorps phospho-SQ ou phospho-TQ) et permit l'identification de 900 sites de phosphorylation[162]. Il faut noter que ce type d'expérience requiert de très grandes quantités de matériel protéique de départ, c'est pourquoi elle n'a pas été utilisée dans cette thèse.

1.7.4.2. Chromatographie d'affinité au métal immobilisé (IMAC)

L'enrichissement en phosphopeptides[163, 164] ou phosphoprotéines[165] par IMAC est une stratégie populaire. Elle a été décrite pour la première fois en 1986 et a permis d'enrichir des isoformes d'ovalbumine phosphorylées[166]. La phase stationnaire IMAC est créée par une réaction de chélation d'un acide iminodiacétique (IDA) ou d'un acide nitrilotriacétique (NTA) avec des ions métalliques triplement chargés. Avec cette approche, les phosphopeptides chargés négativement sont purifiés par affinité à un métal comme Fe^{3+} [167], Ga^{3+} [168] ou Al^{3+} . Cependant, certains phosphopeptides contenant des résidus acides sont également enrichis par cette méthode[169]. Il est possible de bloquer les résidus

acides par O-méthyl estérification afin d'augmenter la spécificité de l'enrichissement. Cependant, la réaction de dérivatisation peut ne pas être complète à 100 % ce qui peut augmenter la complexité de l'échantillon. Au fil du temps, les protocoles n'ont cessé d'être optimisés afin d'améliorer les compositions des tampons de chargement de l'échantillon (0.1-1 % acide acétique ou TFA / 30-60 % ACN), étapes de lavage (0.1-1 % acide acétique / 33 % ACN / 33 % MeOH / 33 % H₂O) et solutions d'élution (0.4 M NH₄OH / 30 % ACN). Il est favorable d'abaisser le pH de l'échantillon afin de protoner les groupes carboxyliques des résidus acides et ainsi augmenter la spécificité[168]. Il a été rapporté que la phase IMAC enrichit préférentiellement les peptides multiples phosphorylés[170, 171] comparativement aux formes mono-phosphorylées.

1.7.4.3. Le dioxyde de titane

Le dioxyde de titane (TiO₂) est une phase stationnaire très efficace dans le but d'enrichir un mélange de peptides en phosphopeptides. En effet, le TiO₂ a une très bonne affinité pour les ions phosphate en solution aqueuse[172]. Il a été décrit par Larsen, *et al.* comme une alternative puissante à la stratégie IMAC[173]. En effet, il a été montré qu'une solution de peptides resuspendue dans une solution acide telle que l'acide 2,5-dihydroxybenzoïque (DHB), l'acide phtalique ou encore l'acide glycolique est significativement enrichie en phosphopeptides suite à son passage à travers une colonne composée de TiO₂ et après élution avec une solution basique (pH 10)[170]. Les conditions d'enrichissement n'ont cessé d'être optimisées et l'acide lactique s'est révélé être une solution de resuspension efficace afin de minimiser la liaison de peptides non-phosphorylés et atteindre un taux d'enrichissement de 90 % en phosphopeptides[174]. En effet, le TiO₂ expose des charges positives à sa surface en condition acide qui vont capturer de manière très efficace les groupements phosphates. De plus, le TiO₂ tolère relativement bien les sels et les traces de détergents ce qui en fait une méthode robuste très appréciée. Contrairement à IMAC, le TiO₂ enrichit de manière préférentielle les peptides mono-phosphorylés comparativement aux peptides multiples phosphorylés. Le mécanisme exact de liaison reste mal compris

mais une explication pourrait être que les peptides multiples phosphorylés se lient également mais sont plus difficilement élués à cause de leur plus grande affinité[171]. C'est la méthode qui a été choisie dans cette thèse à cause de sa très bonne sélectivité et du nombre d'identifications obtenues (tous deux supérieurs à IMAC lors de tests effectués au laboratoire).

Récemment, des nouvelles technologies ont été proposées afin d'effectuer l'enrichissement sur TiO_2 en ligne avec le spectromètre de masse. Par exemple, Agilent a introduit une approche basée sur une puce qui intègre l'enrichissement en phosphopeptides et la séparation par phase inverse en ligne avec le spectromètre de masse[175, 176].

1.7.4.4. L'approche SIMAC (élution séquentielle après IMAC)

L'approche SIMAC est une combinaison intéressante de l'utilisation des phases IMAC et TiO_2 dans le but d'augmenter la couverture du phosphoprotéome et d'identifier les peptides mono- et multiples phosphorylés lors d'une même expérience[171]. Les peptides mono-phosphorylés sont élués par IMAC en condition acide puis les formes multiphosphorylées sont éluées en conditions basiques lors d'une deuxième désorption sur les mêmes billes. La fraction acide est ensuite enrichie par TiO_2 où les phosphopeptides sont élués en conditions basiques. Chacune des fractions est analysée séparément par LC-MS/MS ce qui permet une augmentation considérable du nombre de sites phosphorylés identifiés. Cependant, cet enrichissement comporte plusieurs étapes et le suivi d'un protocole rigoureux est nécessaire afin d'obtenir une bonne reproductibilité et une quantification précise. De plus, le temps d'instrument est deux fois plus long ce qui explique pourquoi le SIMAC n'a pas été retenu pour nos analyses.

Une approche alternative utilise des nanoparticules de diamant recouvertes de résidus poly-arginine et a été testée avec succès d'abord sur un mélange peu complexe de peptides de caséine[177]. Ensuite, cette stratégie a été appliquée en combinaison avec l'utilisation des mêmes nanoparticules recouvertes de TiO_2 dans le but d'identifier les

phosphopeptides mono- et multiplement phosphorylés dans un mélange peu complexe selon le même mode opératoire que SIMAC[178].

1.7.4.5. Capture par affinité aux ions métalliques basés sur un polymère (PolyMAC)

Cette technique innovante a été présentée par Iliuk, *et al.* en 2010 et décrit l'utilisation de dendrimères de polyamidoamine multifonctionnalisés avec des ions titanes et des groupements aldéhydes pour la chélation et l'isolation subséquente des phosphopeptides dans un environnement homogène[179, 180]. Ces dendrimères sont solubles dans l'eau ce qui élimine toute hétérogénéité de liaisons qui pourrait limiter la reproductibilité des analyses. Un schéma de la stratégie PolyMAC est présenté en **Figure A1. 1**. Cette technique, utilisée en combinaison avec un enrichissement grâce à un anticorps anti-phosphotyrosine, a permis l'identification de 794 tyrosines phosphorylées dans des cellules cancéreuses. Elle est rapportée comme une méthode reproductible et très sélective pouvant même surpasser le TiO_2 [179]. Cette approche a été testée au laboratoire et n'a pas donné les résultats escomptés en termes de sélectivité et de nombre total d'identifications en phosphopeptides (données non publiées).

De manière globale, les méthodes d'enrichissement en phosphopeptides basées sur des affinités au métal décrites ci-dessus comportent une étape de conditionnement de la phase suivie du chargement de l'échantillon, du lavage des interactions non-spécifiques et de l'élution des phosphopeptides. Les phases sont le plus souvent conditionnées sous forme de micro-colonnes réalisées dans des cônes de pipette (20 μL ou 200 μL) pourvus d'une phase hydrophobe à son extrémité afin de retenir les particules lors de leur centrifugation.

1.7.4.6. **Autres chromatographies d'affinité pour l'enrichissement en phosphopeptides**

Le dioxyde de zirconium (ZrO_2) est également utilisé pour l'enrichissement en phosphopeptides et a permis l'identification de 248 sites de phosphorylation dans un lysat de foie de souris[181]. L'hydroxyde d'aluminium ($AlOH_3$) a été testé à petite échelle et a été utilisé pour un échantillon de caséine peu complexe[182]. L'oxyde de niobium (Nb_2O_5) a fait l'objet d'une évaluation et a montré un taux d'enrichissement de 50 à 100 % sur un ensemble de phosphopeptides synthétiques[183]. Cependant, ces méthodes ne sont pas comparables aux résultats obtenus avec TiO_2 ou IMAC et ne seront pas décrites en détails. Certaines méthodes impliquant le couplage de Fe^{3+} , Ga^{3+} ou de TiO_2 sur des billes magnétiques a déjà été rapporté pour une meilleure routine d'utilisation et pour simplifier les étapes de lavage[184]. La même stratégie a déjà été appliquée au phosphonate de zirconium[185]. Il existe également des procédures d'enrichissement spécifiques pour une analyse subséquente des phosphopeptides par MALDI qui peuvent s'effectuer directement sur la plaque de dépôt. Ces plaques peuvent être conçues pour un enrichissement par IMAC, oxyde de zirconium ou encore par TiO_2 tout en prenant soin de sélectionner une matrice appropriée[186, 187].

1.7.4.7. **Précipitation au phosphate de calcium**

En 1994, une technique d'enrichissement impliquant la précipitation des phosphopeptides de caséine au phosphate de calcium dans une solution de 50 % éthanol a été utilisée[188]. Plus récemment, cette stratégie a été adaptée avec du phosphate de disodium et du chlorure de calcium en combinaison avec IMAC et a permis d'identifier 227 sites de phosphorylation[189].

1.7.4.8. Dérivatisation chimique

Diverses techniques de dérivation chimiques ont été utilisées au cours du temps afin d'enrichir des échantillons complexes en phosphopeptides. Elles sont basées sur le fait qu'en conditions alcalines, l'élimination β du groupement phosphate sur un peptide phosphorylé sur une sérine ou une thréonine génère respectivement une déhydroalanine et un acide déhydroaminobutyrique. Il est ensuite possible d'ajouter un composé chimique au groupe sulfhydryle libre par addition de type Michael, créant ainsi un groupe di-thiol qui sert d'agent de liaison pour la biotine qui peut être purifiée par des groupements avidine[190] ou par un autre agent de liaison[191]. Un schéma de la réaction est présenté en **Figure A1. 2**. Cette approche a aussi été utilisée dans le but de conduire une protéolyse spécifique des résidus phosphosérines et phosphothréonines convertis respectivement en aminoéthylcystéine et β -aminoéthylcystéine qui sont des cibles pour des protéases spécifiques aux lysines (trypsine, Lys-C...). Ainsi, des fragments peptidiques sont générés au niveau des sites auparavant phosphorylés[192]. Il faut tout de même noter qu'uniquement les diastéréoisomères de type R seront de bons substrats, ce qui signifie qu'uniquement 50 % des sites précédemment phosphorylés sont clivés et possiblement identifiés. De plus, un des inconvénients majeurs de l'élimination β provient de la contamination de peptides portant d'autres modifications telles que la O-glycosylation. Malgré tout, cette stratégie permet d'éviter le problème de mauvaise fragmentation des phosphopeptides et facilite leur identification. Cependant, elle n'est applicable qu'aux phosphopeptides phosphorylés sur des sérines ou thréonines.

La chimie phosphoramidate est une approche qui comprend plusieurs étapes de modifications chimiques mais qui est applicable à tous les résidus phosphorylés. Elle a été introduite en 2001 par Zhou, *et al.* et consiste en la O-méthyl estérification de tous les groupements carboxyliques peptidiques afin de les empêcher de réagir lors des étapes subséquentes ce qui engendrerait une purification non spécifique[193]. Les phosphopeptides sont ensuite dérivés avec un groupement sulfhydryle puis liés à une fonction iodoacétyle immobilisée sur un support de polymère solide[194] ou encore sur des

billes de verre[195]. Ensuite, les groupements phosphate sont reconstitués par hydrolyse acide de la liaison phosphoramidate afin de faciliter leur identification par MS. Un schéma de la réaction est présenté en **Figure A1. 3**. Il faut noter que cette stratégie est particulièrement coûteuse et que les nombreuses étapes impliquées peuvent engendrer des pertes. Elle a cependant été utilisée par Bodenmiller, *et al.* dans le but d'isoler des phosphopeptides dans des cellules Kc167 de drosophile et pour distinguer des sites de phosphorylation précis du facteur de transcription purifié Foxo[195].

Une autre stratégie consiste en une approche réversible sur surface solide basée sur une réaction d'oxydo-réduction et de condensation de phosphopeptides avec une résine de Wang dérivatisée avec des glycines[196].

D'une manière générale, nous n'utilisons pas les méthodes de dérivatisation chimique pour les études phosphoprotéomiques au laboratoire car nous nous intéressons à l'aspect quantitatif des données. Il est probable que la réaction chimique ne soit pas complète ce qui pourrait introduire de la variabilité lors d'analyses comparatives.

Il est important de mentionner qu'au fil des années, l'enrichissement en phosphopeptides a été considéré comme un art à part entière et qu'il convient d'utiliser la technique la plus appropriée au laboratoire. En effet, bien que de nombreux protocoles aient été publiés récemment, tous s'accordent sur le fait que des optimisations doivent être réalisées dépendamment du type cellulaire étudié, du tampon de lyse et de l'enzyme utilisés. De plus, une certaine maîtrise de la technique s'acquiert uniquement avec l'expérience du manipulateur. Au laboratoire, nous avons développé une expertise solide pour l'utilisation du TiO₂, c'est pourquoi elle est notre méthode de choix lors de nos analyses phosphoprotéomiques. Une étape de déssalage est généralement ajoutée après l'enrichissement en phosphopeptides et avant l'analyse par LC-MS/MS.

1.8. Séparation par chromatographie liquide

La chromatographie liquide (LC) est une méthode de séparation de composés chimiques où l'échantillon est entraîné dans une phase mobile le long d'une phase stationnaire. Chaque molécule se déplace à une vitesse différente dépendant de ses propriétés et de son affinité vis-à-vis de la colonne. Il est possible de séparer des protéines intactes ou des peptides par LC. Dans cette thèse, nous avons décidé de fractionner l'échantillon au niveau peptidique en ligne avec le spectromètre de masse ce qui nous permet de mieux contrôler la routine d'injection et d'avoir une reproductibilité optimale.

1.8.1. Séparation selon l'hydrophobicité

La séparation de peptides sur phase inverse est la stratégie la plus utilisée en protéomique. La phase non polaire est constituée de chaînes alkyles liées de manière covalente à une matrice de silice en général. La stratégie la plus populaire est l'utilisation de chaînes de 18 ou 8 atomes de carbones (C_{18} ou C_8). Les peptides éluent de la colonne grâce à une phase mobile dont la composition en solvant organique augmente au cours du gradient. Ainsi, les peptides sont désorbés proportionnellement à leur hydrophobicité. Des peptides très hydrophiles sont peu ou pas retenus sur la colonne tandis que des peptides très hydrophobes ont de la difficulté à être élués. Ce type de chromatographie est communément utilisé en ligne avec la MS.

1.8.2. Séparation par échange d'ions

L'utilisation de phases stationnaires échangeuses d'ions est très efficace pour séparer des protéines ou des peptides selon leurs charges. Les matériaux polymériques utilisés sont constitués soit de groupements fonctionnels anioniques pour une chromatographie par échange fort de cations (SCX) ou bien cationiques pour un échange fort d'anions (SAX).

Ces groupements sont attachés de manière covalente à une résine constituée de billes de cellulose ou d'agarose. Les interactions ioniques mises en jeu régulent la rétention des peptides sur les différentes phases. La séparation de phosphopeptides par chromatographie échangeuse d'ions est utilisée comme méthode de préfractionnement avant ou après une étape d'enrichissement mais n'est pas considérée à elle seule comme très efficace dans le but unique d'isoler des phosphopeptides dans un mélange complexe. Dans cette thèse, nous avons utilisé une séparation par échange de cations où les peptides chargés positivement circulent dans la colonne avec des vitesses différentes. A l'inverse, les anions sont très peu retenus du fait qu'il ne se créera pas d'interaction avec la phase. L'élution des ions positifs retenus se fait par compétition avec une solution de sels contenant des ions positivement chargés. Les phosphopeptides ont en théorie une affinité de liaison moindre avec la phase SCX à cause de leur charge négative et devraient alors éluer avant les peptides non phosphorylés qui ont une charge globale plus positive en condition acide.

Une étude menée par Beausoleil, *et al.* a démontré l'intérêt de cette stratégie dans le but d'étudier le phosphoprotéome nucléaire de cellules Hela où 2002 sites de phosphorylation ont été identifiés[197]. Cependant, de nombreuses études s'accordent sur le fait que plusieurs paramètres entrent en compte lors de la séparation de phosphopeptides par SCX et qu'en général leur élution est répartie à travers toutes les fractions et n'est pas concentrée uniquement dans les premières[161, 198]. En effet, la proportion de résidus positivement chargés tels que les lysines, arginines ou histidines ainsi que la charge portée par le N-ter contrebalancent les charges négatives provenant du phosphate.

Il est possible de combiner deux types de chromatographies l'une à la suite de l'autre afin de rendre l'échantillon moins complexe et d'augmenter le nombre d'identifications d'une manière considérable lors de l'analyse par MS. Cette stratégie appelée MudPIT (technologie d'identification de protéine en plusieurs dimensions) peut être réalisée de façon séparée en mode « hors ligne » ou « en ligne » dépendamment si le fractionnement est couplé ou non au spectromètre de masse. Cette approche est utilisée

dans de nombreuses publications et permet d'obtenir une plus grande couverture du phosphoprotéome dans divers modèles expérimentaux tels que des cellules Hela humaines[19], IEC-6 de rat[199] ou encore S2 chez la drosophile[1]. Il faut cependant noter que la stratégie MudPIT nécessite une grande quantité de matériel de départ et que les méthodes de quantification utilisées en aval doivent être robustes. En effet, les phosphopeptides éluent dans plusieurs fractions consécutives en général et chacune de leur intensité doit être répertoriée pour la quantification.

Une étude comparative a été publiée en 2011 dans le but de comparer les méthodes chromatographiques décrites précédemment et a démontré la supériorité de la combinaison SCX/C₁₈ lors d'études phosphoprotéomiques[200]. A ce jour, celle-ci reste la méthode de choix pour de nombreux groupes de recherche car elle constitue une stratégie puissante et nous l'utilisons en routine au laboratoire.

1.9. La spectrométrie de masse comme outil d'étude de la phosphorylation

La spectrométrie de masse est la méthode de choix afin d'identifier et quantifier des événements de phosphorylation protéiques tout en permettant la localisation précise du site modifié dans la séquence[201]. Le spectromètre de masse génère un spectre représentant l'intensité du signal d'un ion selon son ratio masse sur charge (m/z).

Généralement, deux approches sont possibles lors d'analyses protéomiques, l'approche ascendante (bottom-up) et la stratégie descendante (top-down). La première consiste à digérer des protéines de toutes tailles en vue de l'analyse de leurs peptides afin de les identifier. Il est ensuite possible de détecter les m/z des peptides présents dans le mélange et de les identifier par MS/MS en plus de localiser les modifications post-traductionnelles avec précision. L'approche descendante consiste en l'injection directe de petites protéines (des histones par exemple) sous leur forme intacte dans le spectromètre de

masse. Il est possible de les fragmenter ensuite afin de les identifier. Cette méthode a pour avantage d'avoir une vue d'ensemble de toutes les modifications présentes sur la protéine en fonction de la masse globale calculée mais leur localisation est moins précise. C'est l'approche ascendante qui a été choisie dans cette thèse afin d'étudier la phosphorylation à grande échelle.

1.9.1. Composantes d'un spectromètre de masse

Un spectromètre de masse est composé de trois parties : une source d'ionisation, un analyseur et un détecteur. L'échantillon est introduit dans l'instrument suite à son ionisation puis les ions formés sont séparés selon leur m/z dans l'analyseur. Finalement, le détecteur convertit l'énergie des ions en signaux électriques transmis à un ordinateur.

1.9.1.1. Sources d'ionisation

L'échantillon peut être introduit dans le spectromètre de diverses manières. Il peut provenir directement d'un système de chromatographie liquide ou gazeuse en ligne avec l'instrument, il peut être infusé sans séparation ou encore être déposé sur une plaque. Les deux modes les plus utilisées en protéomique sont l'ionisation par électrospray (ESI) à pression atmosphérique et la désorption/ionisation assistée par matrice (MALDI) sous vide. Les modes APCI (Ionisation Chimique à Pression Atmosphérique)[202], FAB (bombardement par atome rapide)[203], DESI (Ionisation par Désorption/Électrospray)[204], SIMS (spectrométrie de masse d'ion secondaire)[205], PDMS (spectrométrie de masse de désorption au plasma)[206] ainsi que les procédés d'ionisation chimique ou par électrons ne sont pas expliqués en détails ici.

1.9.1.1.1. *Électronébulisation (ESI)*

L'ionisation par électronébulisation est réalisable en mode positif ou négatif. Pour des peptides en condition acide, le mode positif est préférable. Lors de l'ionisation d'un peptide, un haut voltage (environ 3 000 V) produit un gradient électrique qui sépare les charges à la surface du fluide à l'extrémité d'un émetteur. Les molécules ionisées en solution induisent une pulvérisation de gouttelettes hautement chargées. Ceci crée la formation d'un cône de Taylor à la sortie jusqu'à ce que le liquide atteigne la limite de Rayleigh où la tension de surface et les répulsions électrostatiques sont égales. Les gouttelettes éjectées sont captées à l'entrée du spectromètre puis des répulsions coulombiennes font encore augmenter la tension de surface. Ceci cause des explosions sous forme de gouttelettes plus petites suite à l'évaporation du solvant qui relargue les ions[207, 208]. Ce mode d'ionisation permet d'analyser des peptides ou des protéines (jusqu'à plusieurs MDa dans le cas de virus entiers par exemple[209]), permet une sensibilité de l'ordre du femtomole, est compatible avec une séparation LC en amont, permet d'obtenir des ions multiples chargés et n'induit pas d'interférence avec une matrice. C'est la méthode de choix que nous utilisons au laboratoire dans un mode nano-ESI à un flux de 600 nL/min.

Lors de l'ionisation par électronébulisation en mode positif, les peptides tryptiques sont protonés en condition acide à cause de la présence du groupement amine en N-ter et d'un acide aminé basique en C-ter (K ou R). Ils sont en général doublement chargés mais leur état de charge peut augmenter dépendamment de la présence d'autres acides aminés basiques comme les histidines dans leur séquence peptidique. Il est possible de déterminer l'état de charge z d'un peptide selon le profil isotopique des pics associés à un certain m/z . Ces pics sont séparés d'une masse de 1 Da, ce qui est causé majoritairement par la probabilité de 1.11 % pour chaque carbone d'être un isotope ^{13}C plutôt que ^{12}C (on estime la présence négligeable de ^{14}C à 10^{-10} %). La probabilité de détecter les isotopes ^2H , ^{15}N ou ^{17}O plutôt que ^1H , ^{14}N ou ^{16}O contribue au profil isotopique de manière minoritaire. La charge d'un peptide est définie selon son profil sur un spectre MS en observant

l'espacement entre 2 pics adjacents qui correspond à la valeur $1/z$. Ainsi, si la différence entre 2 pics est de 0.5 unités, le peptide sera doublement chargé comme présenté dans l'exemple de la **Figure 1.8**.

1.9.1.1.2. *Ionisation par MALDI (désorption/ionisation assistée par matrice)*

L'analyse par MALDI consiste en l'ionisation d'un échantillon suite à l'excitation et l'ablation de la matrice dans laquelle il est resuspendu. L'analyte est co-cristallisé dans un excès de matrice puis le mélange est irradié par un pulse laser qui engendre sa vaporisation. Bien que le mécanisme exact ne soit pas complètement élucidé, on pense que l'ionisation de l'échantillon se fait suite à un transfert de proton de la matrice à l'analyte pendant la désorption. Une fois en phase gazeuse, les molécules sont dirigées vers l'analyseur qui est un tube de temps de vol (TOF) en général. Cette méthode permet d'analyser des ions jusqu'à plusieurs MDa, a une bonne sensibilité de l'ordre du femtomol mais la matrice peut causer des interférences lors de la détection de molécules de faibles masses (< 1 000 Da). La composition de la matrice peut être ajustée en fonction de l'échantillon étudié (acide α -cyano-4-hydroxycinnamique, acide sinapinique, acide 2,5-dihydroxybenzoïque...).

1.9.1.2. **Analyseurs de masse**

Parallèlement aux avancées technologiques dans le domaine des sources d'ionisation, les performances des analyseurs n'ont cessé de s'améliorer en termes de vitesse, précision et résolution. La vitesse est déterminée par le temps nécessaire au spectromètre de masse pour compléter un cycle de scans MS et MS/MS afin de séquencer un maximum de précurseurs distincts. La précision reflète l'habileté de l'analyseur à obtenir un ratio m/z qui soit le plus proche de sa valeur théorique. Cette valeur est souvent décrite en ppm (partie par million).

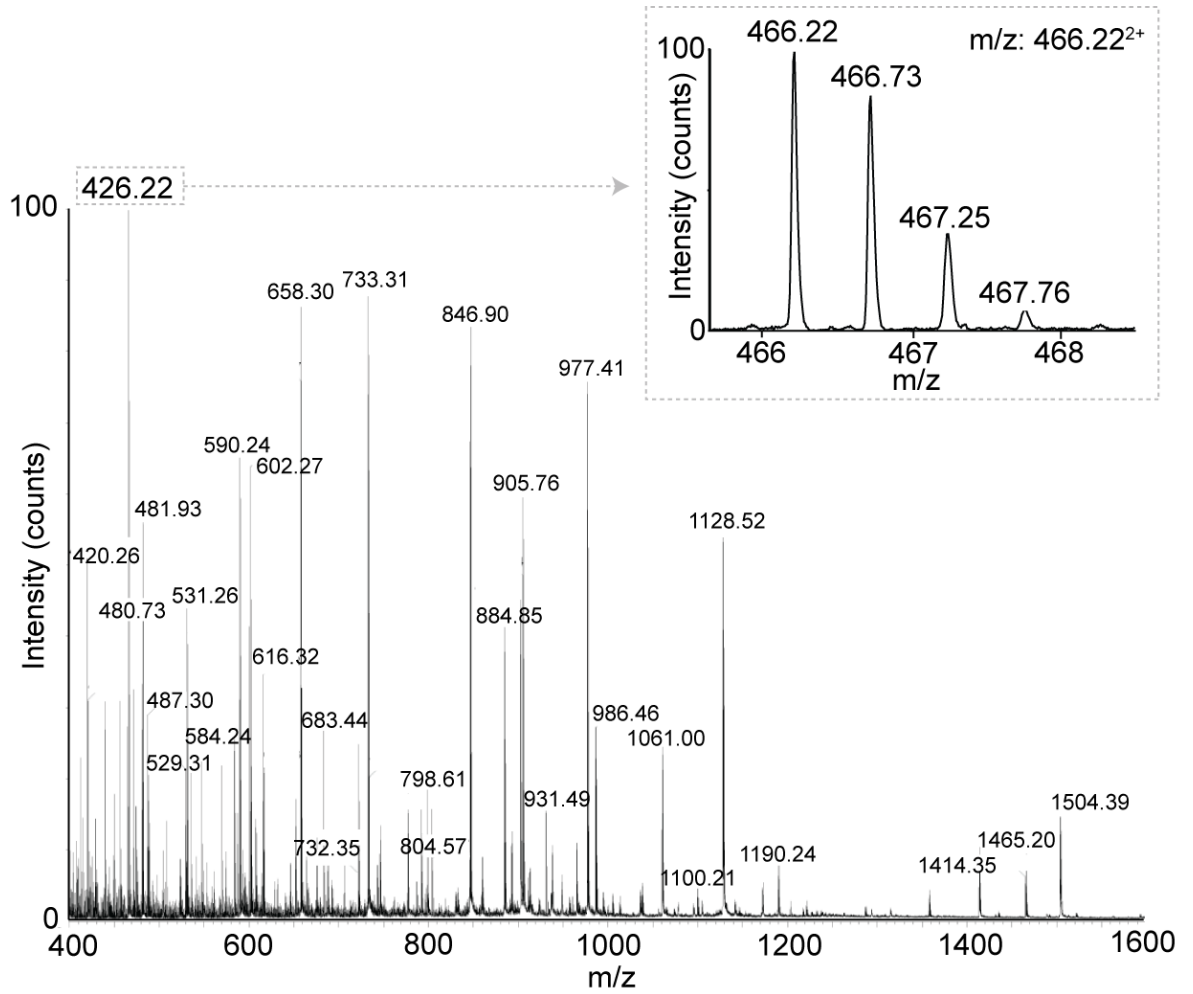


Figure 1.8. Spectre de masse d'un mélange de peptides représentant leur intensité en fonction de leur ratio masse sur charge (m/z).

Le profil isotopique présenté en encadré montre un espacement de 0.5 unités entre deux pics adjacents et confirme l'état doublement chargé du peptide ayant un m/z : 466.22²⁺.

La résolution correspond à la capacité de distinguer des ions de différents ratios m/z . Elle est donnée par l'équation : $résolution = M/\Delta M$ où M correspond au m/z et ΔM représente la largeur du pic à mi-hauteur. Le pouvoir de résolution est un facteur déterminant pour que l'instrument puisse distinguer les différents pics isotopiques et non un large pic unique lors de l'analyse du précurseur.

1.9.1.2.1. *Quadrupôles*

Un quadrupôle est constitué de quatre barres cylindriques parallèles entre elles. Chaque paire de quadrupôles diamétralement opposée est connectée directement à un générateur de fréquences radio (RF) qui permet uniquement aux ions ayant un ratio m/z spécifique d'être transmis à travers le dispositif. Un courant continu est superposé aux ondes radio et c'est la stabilité différentielle de la trajectoire des ions qui permet de les séparer. Certains ions sont transmis alors que d'autres entrent en collision avec l'une ou l'autre des barres et seront neutralisés. Il suffit de scanner le champ de fréquences radio afin de détecter les ions dans un intervalle de 400 à 1600 m/z pour des peptides en général. La résolution est de 4 000 et la précision de masse de 100 ppm. Lors d'une analyse par spectrométrie de masse en tandem, il est possible d'aligner trois quadrupôles à la suite. Le premier servira à scanner l'intervalle de m/z , le deuxième jouera le rôle de cellule de collision où un gaz sera introduit en vue de la fragmentation d'ions spécifiques et le troisième permet d'analyser les ions fragments.

1.9.1.2.2. *Trappe ionique*

Il existe deux types de trappes ioniques : la trappe 3D et la trappe linéaire. Dans une trappe 3D (ou trappe de Paul), les ions sont accumulés à l'intérieur du dispositif qui comporte deux électrodes terminales et une électrode centrale. Les ions ont une trajectoire stable en appliquant un voltage RF sur l'électrode centrale. En augmentant le voltage RF ou en

appliquant un voltage auxiliaire aux électrodes, il est possible de déstabiliser les ions et de les éjecter. On peut aussi forcer l'accumulation d'un ion et induire sa fragmentation suite à l'application d'une longueur d'onde spécifique d'excitation au niveau des électrodes terminales.

La trappe ionique linéaire (LIT) ou trappe linéaire quadrupolaire (LTQ) est constituée de barres formant un quadrupôle contenant les ions dans un champ électrique en deux dimensions grâce à des fréquences radio et en appliquant un potentiel aux extrémités des électrodes. Elle peut être utilisée comme un filtre de masse et servir d'enceinte de fragmentation. Elle est remplie par un faisceau ionique continu et les ions piégés sont accumulés puis séparés de manière plus efficace que dans une trappe 3D. La résolution est assez faible ($\sim 4\ 000$) et la précision de masse est de l'ordre de 100 ppm.

1.9.1.2.3. *Orbitrap*

L'orbitrap, inventée par Alexander Makarov, consiste en deux électrodes. L'électrode externe a la forme d'un baril tandis que celle située à l'intérieur a la forme d'un fuseau coaxial qui crée un champ électrostatique avec une distribution potentielle quadrol logarithmique (combinaison du champ quadrupolaire d'une trappe ionique et du champ logarithmique d'un condensateur cylindrique)[210]. L'équation associée à ce champ $U(r, z)$ est :

$$U(r,z) = \frac{k}{2} \left(z^2 - \frac{r^2}{2} \right) + \frac{k}{2} (R_m)^2 \ln \left[\frac{r}{R_m} \right] + C$$

Équation 1.

où r et z sont les coordonnées cylindriques, C est une constante, k est la courbure du champ et R_m est le rayon caractéristique.

Les ions sont injectés de manière tangentielle au champ électrique appliqué entre les deux électrodes et sont piégés à cause du fait que leur attraction électrostatique vis-à-vis de l'électrode interne est balancée par des forces centrifuges. Ainsi, les ions oscillent autour de l'électrode centrale et ont également un mouvement de va-et-vient latéral, créant un déplacement global de spirale complexe. Les ions avec un m/z spécifique ont une fréquence d'oscillation harmonique qui leur est propre, indépendante de leur vitesse et inversement proportionnelle à la racine carrée de leur m/z . Les oscillations axiales sont détectées par signaux de courant et sont converties en spectre de masse par transformée de Fourier (FT). L'équation associée à la fréquence des oscillations axiales ω (rad/s) est :

$$\omega = \sqrt{(q/m)k}$$

Équation 2.

où q est la charge de l'ion, m est sa masse et k est la courbure du champ.

Les ions analysés dans une orbitrap ont une très haute précision de masse (< 2 ppm) et une résolution pouvant atteindre 200 000.

1.9.1.2.4. Tube de temps de vol (TOF)

L'analyseur de type tube de temps de vol est basé sur l'accélération cinétique d'un groupe d'ions vers un détecteur où tous les ions de même charge acquièrent la même énergie. Comme leurs m/z sont différents, les ions ayant le m/z le plus petit atteindront le détecteur en premier tandis que les peptides avec un plus haut m/z prendront plus de temps pour s'y rendre. Il est possible d'améliorer la résolution en ajoutant un réflectron qui va augmenter le temps requis pour les ions afin d'atteindre le détecteur et va diminuer leur distribution d'énergie cinétique. La résolution peut alors atteindre 40 000 et la précision de masse de 5-10 ppm.

Il est possible de combiner différents analyseurs l'un à la suite de l'autre. C'est le cas notamment des instruments tels que le quadropôle-tube de temps de vol (Q-TOF) ou encore la trappe ionique linéaire-orbitrap (LTQ-Orbitrap). Les spectromètres de masse hybrides utilisés dans cette thèse sont le LTQ-Orbitrap XL, l'Orbitrap Velos et l'Orbitrap Elite commercialisés par Thermo Scientific. L'Orbitrap Elite est l'instrument le plus performant à ce jour pour les études protéomiques et a été décrit dans une première publication début 2012 par le groupe de M. Mann[211]. L'Orbitrap Elite comprend deux trappes linéaires Velos PRO (une à haute et une à basse pression) pour la transmission et la fragmentation en mode CID et permet d'obtenir une très haute résolution sur les précurseurs dans l'orbitrap ($> 240\ 000$). Les modes de fragmentation HCD et ETD sont également implémentés dans l'instrument. Un schéma détaillé de cet instrument est présenté en **Figure 1.9. A**. Il comprend une nouvelle génération d'optiques pour la transmission optimale des ions qui inclut notamment des lentilles S sous haute pression afin de concentrer les ions de manière très efficace. Un nouveau quadropôle courbé permet de guider les ions provenant d'un octopôle (**Figure 1.9. B**). L'analyse de masse dans la trappe ionique est très rapide ($66\ 000$ amu/s) et permet d'effectuer 12.5 MS/MS par seconde en mode CID. La gamme dynamique de détection est également améliorée (jusqu'à six ordres de grandeur) grâce à l'utilisation de détecteurs sous forme de dynodes pour la multiplication des électrons. Cet atout permet d'obtenir une meilleure précision lors d'études quantitatives et d'améliorer la limite de détection. De plus, une orbitrap plus compacte a été installée en combinaison avec de nouvelles lentilles miniatures ce qui permet d'augmenter nettement la sensibilité (**Figure 1.9. C**). Le mode HCD permet d'analyser les peptides en stratégie « haute-haute » où les précurseurs et les ions fragments sont analysés à haute résolution dans l'orbitrap[211].

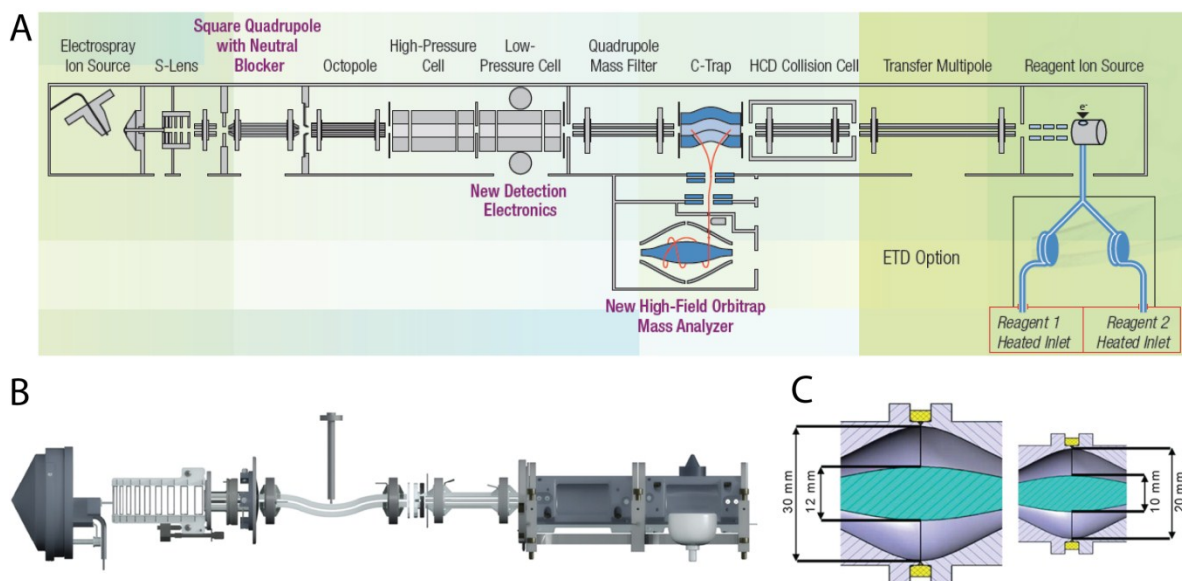


Figure 1.9. Spectromètre de masse Orbitrap Elite commercialisé par Thermo Scientific.

A. Schéma détaillé représentant toutes les composantes du spectromètre incluant les options de fragmentations ETD et HCD. B. Représentation des nouvelles générations d'optiques pour le transfert des ions incluant la lentille S à gauche, un quadrupôle courbé puis la double trappe ionique à droite. C. Comparaison des dimensions de l'orbitrap standard retrouvée sur le Velos ou le XL (gauche) avec celle installée sur le Elite (droite) qui est plus compacte. Adapté de [211] avec la permission de American Society for Biochemistry and Molecular Biology, Molecular & Cellular Proteomics, copyright 2012.

1.9.1.3. Détecteurs

Le détecteur le plus utilisé est le multiplicateur d'électrons (composé de dynodes d'oxyde d'aluminium) qui transfère l'énergie cinétique des ions rentrant en collision avec sa surface, ce qui génère une cascade d'électrons secondaires convertis en courant. La coupe de Faraday (moins sensible et qui offre un niveau d'amplification limité) implique des dynodes de BeO, GaP, ou CsSb et génère également l'émission d'électrons secondaires. La dynode de conversion photomultiplicatrice utilise un écran constitué de phosphore au lieu

d'une dynode. Au contact des ions, elle génère l'émission de photons détectés sur une surface photomultiplicatrice qui émet des électrons amplifiés en cascade. Le détecteur à puces consiste en plusieurs détecteurs individuels alignés qui vont détecter des ions de différents m/z séparés dans l'espace.

1.9.2. Spectrométrie de mobilité ionique modulée par un champ électrique en forme d'onde asymétrique à champ élevé (FAIMS)

FAIMS est une technologie capable de séparer des ions en phase gazeuse à pression atmosphérique (760 torr). Les précurseurs de cette technique ont été développés à l'origine par une équipe russe dirigée par M.P. Gorshkov dans les années 1980[212]. Par la suite, Roger Guevremont et Randy W. Purves ont publié leurs travaux et décrit la première électrode cylindrique FAIMS en 1998[213]. FAIMS fait partie des méthodes de séparation par mobilité ionique au même titre que DTIMS (spectrométrie par mobilité ionique en fonction du temps de dérive), AIMS (spectrométrie par mobilité ionique d'aspiration) et TWIMS (spectrométrie de mobilité ionique par onde progressive) qui ne seront pas décrites en détails. C'est une méthode orthogonale à la chromatographie liquide et à la spectrométrie de masse. Le dispositif FAIMS est installé à la sortie de la colonne de chromatographie liquide et devant l'entrée du spectromètre de masse. Contrairement à l'utilisation du mode en électronébulisation conventionnel qui transmet tous les ions vers le spectromètre de masse (**Figure 1.10. A**), le FAIMS permet de sélectionner la population transmise (**Figure 1.10. B**).

Les ions sont transportés dans un flux de gaz puis séparés selon leur différence de mobilité ionique dans un champ électrique élevé. Lorsqu'un champ électrique créé par l'application d'une onde asymétrique est faible (< 200 volts/cm), l'ion a un mouvement de va et vient ou oscille sur place mais ne touche jamais les électrodes. Par contre, si le champ est élevé ($\sim 5\,000$ volts/cm), alors l'effet de l'onde incite l'ion à aller dans l'une ou l'autre des directions.

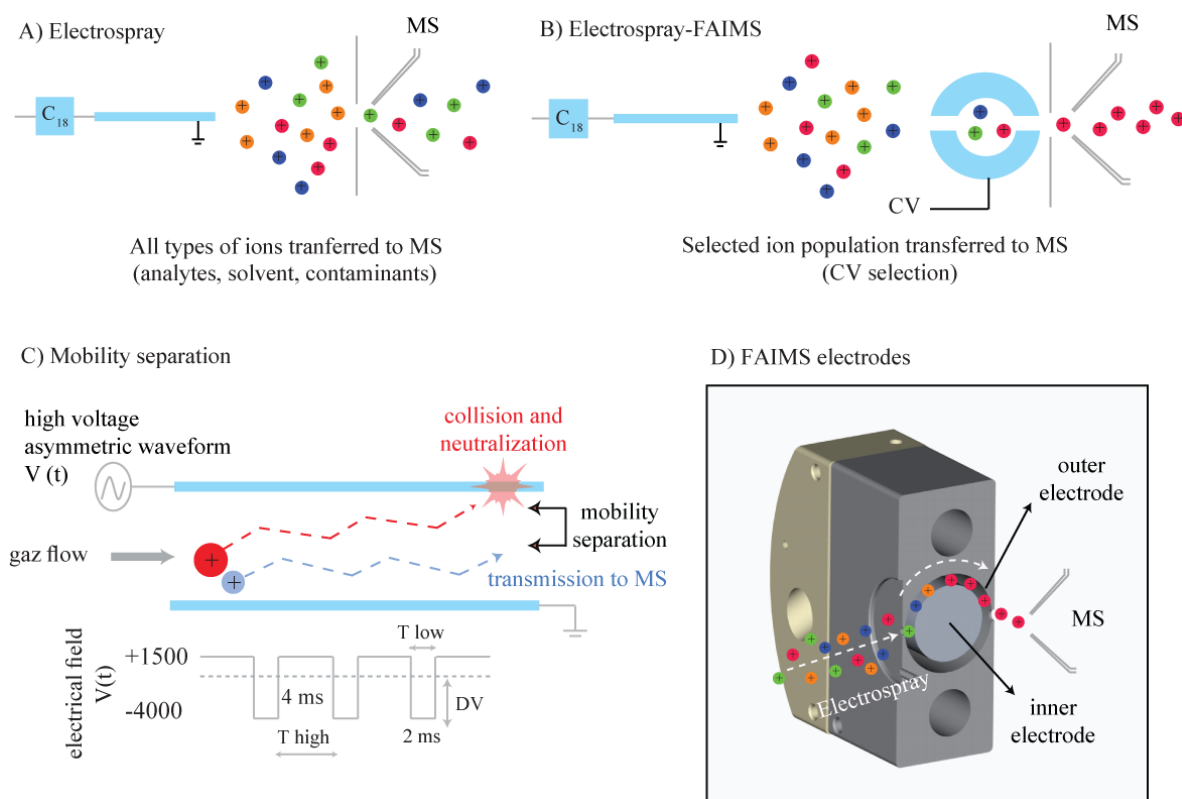


Figure 1.10. Représentation de la séparation d'ions par FAIMS.

A. L'électronébulisation conventionnelle permet à tous les ions élués de la colonne d'être transmis vers le spectromètre de masse. B. Avec FAIMS, seulement une partie des ions est transmise lors de l'application d'un voltage de compensation (CV) spécifique. C. Représentation schématique de la séparation de deux ions par FAIMS selon leur mobilité ionique différentielle dans un champ électrique élevé. L'ion rouge va entrer en collision avec une électrode et ne sera pas transmis vers le spectromètre de masse tandis que le voltage de compensation appliqué sera favorable à la transmission de l'ion bleu. D. Schéma des électrodes concentriques circulaires (une interne et une externe) composant le dispositif FAIMS.

L'ion doit être dans une condition « balancée » afin d'être transmis à travers le FAIMS. La différence de mobilité ionique d'un ion dans un champ électrique faible et élevé est telle que sa vitesse est plus importante lors de l'application du champ élevé. Ainsi, la distance parcourue par l'ion dans ce champ est légèrement plus grande. Étant donné que l'onde est appliquée à haute fréquence (> 200 kHz), la somme de chaque petite différence de distance supplémentaire parcourue dans chaque période résulte en une différence significative à la

fin du parcours. L'onde électrique asymétrique est appliquée entre deux surfaces conductrices espacées de quelques mm. L'asymétrie de l'onde consiste en une différence significative entre l'intensité du plus haut voltage et celle du plus bas voltage. Elle est communément de forme rectangulaire et appliquée pendant une courte période de temps en voltage positif puis en voltage négatif pour une plus longue durée. Cependant, les aires des portions positive et négative restent égales afin de ne pas influencer la séparation. La différence de voltage est appelée voltage de dispersion (DV) et est de l'ordre de -5 000 V en général (**Figure 1.10. C**).

La dérive d'un ion vers une des plaques peut être stoppée par l'application d'un voltage de compensation (CV). Si le voltage est appliqué avec une magnitude et une polarité appropriée, la trajectoire de l'ion est détournée dans le champ. Lors de l'application d'un CV précis, la plupart des ions dans un mélange va toucher l'une ou l'autre des plaques tandis qu'une certaine proportion d'entre eux est dans un état balancé et est transmise à travers le dispositif. Un balayage de CV peut être réalisé avec des sauts de voltage plus ou moins espacés afin de générer un spectre de CVs et déterminer lequel permet la transmission maximale d'un ion spécifique. Le gaz utilisé doit permettre aux ions d'atteindre rapidement leur vitesse finale qui est proportionnelle à la force du champ appliqué. Cette proportionnalité est spécifique à chaque analyte et permet de séparer des composés dans un mélange complexe. La séparation des ions s'effectue selon leur taille et leur forme globale (« cross section »).

Ainsi, il est possible de distinguer des isomères, des conformères ou des ions isobariques qui ont le même m/z et co-éluent en LC mais qui n'auront pas exactement la même structure tridimensionnelle et donc seront potentiellement séparables selon leur différence de mobilité ionique. Cette application a déjà été démontrée sur des peptides synthétiques[214-218]. C'est le cas notamment pour les phosphopeptides isomériques ou bien souvent la localisation du site de phosphorylation reste ambiguë à cause de leur co-élution et de la génération d'un spectre MS/MS mixte. Nous avons démontré cette application sur des phosphopeptides provenant d'un extrait protéique de cellules S2 de

drosophile[1] et cela a été également démontré avec des phosphopeptides synthétiques[218].

Deux types de dispositifs FAIMS existent et sont constitués soit d'électrodes planaires, soit d'électrodes concentriques cylindriques. Dans le cas d'électrodes cylindriques (une interne et une externe), les ions sortent de manière radiale fixe, c'est-à-dire qu'ils sont concentrés à une position spécifique entre les électrodes (**Figure 1.10. D**). Leur distribution est nette à cause de la diffusion minimale, de la répulsion très faible entre ions et du mouvement de turbulence du gaz. Le flux d'ions est comprimé en volume et sa densité augmente dans le FAIMS tandis que deux forces (la diffusion et la répulsion entre ions) ont tendance à lui faire prendre de l'expansion. Dans le cas d'électrodes planaires, les ions sont repartis de manière uniforme entre les deux plaques ce qui peut provoquer des pertes à cause de la diffusion plus importante. Cependant, la résolution est augmentée lors de l'utilisation de ce dispositif et permet même la distinction entre une leucine et une isoleucine[219-222].

Il faut noter que l'espacement entre les électrodes ainsi que la forme et la fréquence de l'onde asymétrique peuvent être ajustés en fonction des applications[223-225]. La composition du gaz (hélium, azote, oxygène, dioxyde de carbone, hydrogène...) et la proportion de chacun des gaz dans le cas d'un mélange (nous utilisons hélium : azote, 50 : 50) est également essentielle[226-228]. Le flux de gaz utilisé habituellement est de 3.5 L/min. Ceci représente une quantité importante de gaz consommé pour une expérience (soit 2 500 L d'hélium par jour s'il est dilué de moitié avec un autre gaz). Des optimisations concernant la géométrie des électrodes ont été effectuées afin de remplacer le mélange d'hélium par de l'azote pur plus accessible et moins dispendieux[229]. La température de chacune des électrodes diffère d'environ 20°C pour des électrodes circulaires (70°C pour l'électrode interne et 90°C pour l'électrode externe).

Globalement, la résolution du FAIMS est plutôt faible. Dans un dispositif FAIMS avec des électrodes cylindriques, la largeur du pic va être proportionnelle au voltage de dispersion et au CV appliqué. La résolution de l'instrument FAIMS commercialisé par

Thermo Scientific qui a été utilisé dans cette thèse est de 15. Dans le cas d'électrodes planaires, la largeur du pic reste plus constante indépendamment du voltage ce qui augmente la résolution (de l'ordre de 100).

Le FAIMS a été employé pour la première fois en 2002 pour l'analyse de peptides tryptiques dans un mélange peu complexe et a été utilisé avec succès notamment en ce qui concerne la réduction de bruit de fond[230, 231]. Le FAIMS permet de décomplexifier l'échantillon avant son entrée dans le spectromètre de masse afin de maximiser le nombre d'identifications par MS/MS[1, 232-234]. De plus, il permet de filtrer les ions mono-chargés qui ne sont pas transmis dans la même gamme de CVs comparativement aux peptides multiplément chargés. Ainsi, il est possible d'augmenter significativement le ratio signal sur bruit et obtenir une meilleure sensibilité. Habituellement, les peptides sont transmis dans un gamme de CVs allant de -50 V à -10 V. Cependant, lors d'expérience de routine, des sauts de CVs de 4 V sont utilisés afin de balayer l'échantillon tout en minimisant le nombre d'injections. Il est alors possible de ne pas détecter certains peptides qui ne seront pas transmis de manière optimale aux CVs choisis arbitrairement. Il est important de mentionner que la technologie FAIMS est encore relativement récente et des améliorations constantes concernant notamment la miniaturisation de l'instrumentation sont en cours[235].

Le FAIMS n'est pas un dispositif installé de manière permanente sur les instruments de Thermo Scientific (LTQ-Orbitrap XL, Velos ou Elite) et nécessite une mise au point pour chaque expérience. Un instrument récent, le Synapt G2 HDMS de Waters combine la spectrométrie à mobilité ionique à haute efficacité intégrée avec la spectrométrie de masse en tandem à haute résolution en une seule unité[236].

1.9.3. Spectrométrie de masse en tandem : méthodes de fragmentation

Auparavant, la dégradation d'Edman était la seule stratégie en vue d'obtenir une séquence peptidique et consistait à marquer le résidu N-ter avec du phénylthiocyanate puis à le

cliver spécifiquement en condition acide[237]. Cette étape laborieuse devait être répétée plusieurs fois afin d'espérer obtenir la séquence totale du peptide, ce qui nécessitait plusieurs heures d'analyses.

La spectrométrie de masse en tandem permet désormais d'isoler un peptide spécifique et de le fragmenter afin d'obtenir sa séquence en quelques millisecondes. Le spectre représentant les ions fragments est appelé MS/MS ou MS² et est le résultat de la fragmentation d'un ion précurseur au niveau des liaisons entre les acides aminés. D'une manière générale, un mode DDA (acquisition dépendante des données) est implémenté et l'acquisition se fait selon un cycle comprenant un scan MS suivi de plusieurs spectres MS/MS (jusqu'à 20 MS/MS consécutifs pour l'Orbitrap Elite) pour ensuite retourner à un balayage MS. Il est également possible de réaliser une étude plus ciblée grâce à l'utilisation de listes d'inclusion ce qui permet de fragmenter uniquement les peptides d'intérêt.

1.9.3.1. CID (dissociation induite par collision)

La fragmentation CID (dissociation induite par collision) ou CAD (dissociation activée par collision) est réalisée suite à la collision du peptide avec un gaz inerte tel que de l'azote, de l'argon ou de l'hélium dans une cellule de collision. Les ions peptidiques fragmentés par CID produisent majoritairement des ions de type b (si la charge est localisée sur le fragment N-ter) et y (si la charge est localisée sur le fragment C-ter) à partir desquels la séquence du peptide peut être interprétée[238, 239]. Un schéma de la fragmentation peptidique générale est représenté en **Figure 1.11. A** et les ions générés par CID sont indiqués en **Figure 1.11. B**. Chaque fragment d'une série (b ou y par exemple) diffère de son voisin par la masse d'un acide aminé. Ainsi, il est possible en principe de déterminer la séquence complète en acides aminés d'un peptide en considérant uniquement les différences de masse entre chaque pic (**Figure 1.12**). Cependant, la difficulté réside dans le fait que l'information n'est pas toujours complète et qu'il faut réussir à distinguer l'appartenance de chaque pic à la bonne série pour éviter toute confusion.

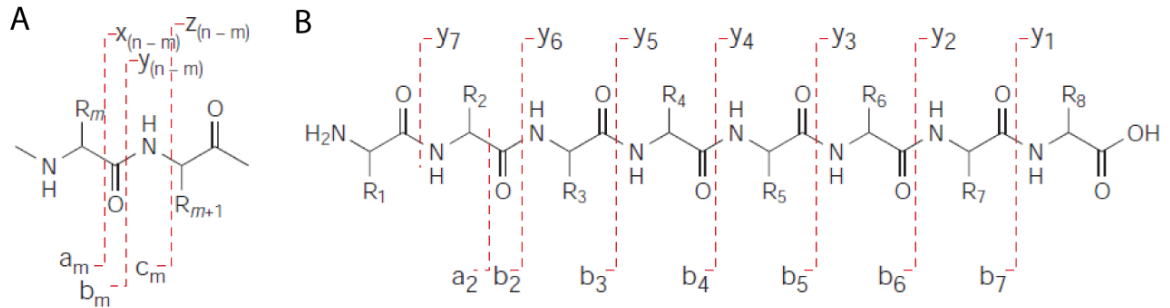


Figure 1.11. Structure chimique d'un peptide et désignation des ions fragments selon la nomenclature Roepstorff–Fohlmann–Biemann suite à une fragmentation.

A. La fragmentation peptidique engendre le clivage de liaisons spécifiques selon le mode utilisé. Cela donne lieu à la formation d'ions $x_{(n-m)}$, $y_{(n-m)}$, $z_{(n-m)}$, a_m^- , b_m^- et c_m^- qui désignent les fragments générés en MS/MS (m représente le nombre de chaînes latérales d'acide aminés contenues dans cet ion et n nomme le nombre total de chaînes latérales contenues dans le peptide). B. La fragmentation CID donne lieu majoritairement à la formation d'ions b et y . Adapté de [240] avec la permission de Nature Publishing Group, Nature Reviews Molecular Cell Biology, copyright 2004.

Quotidiennement, des algorithmes sophistiqués nous aident à obtenir les séquences de manière automatisée. Il est essentiel d'avoir une précision de masse exacte sur le précurseur et sur les fragments afin d'assigner les séquences de la manière la plus fiable possible lors de l'interrogation de bases données. Le traitement bioinformatique des données brutes est décrit dans la **section 1.10**.

Concernant les phosphopeptides, la perte d'acide phosphorique H_3PO_4 en phase gazeuse suite à la fragmentation de séquences contenant des résidus pSer (phosphosérine) et pThr (phosphothréonine) est souvent observée comme un pic majoritaire lors de la fragmentation CID. Le réarrangement consécutif à l'élimination de type β de la liaison phosphodiester sur les pSer et pThr génère respectivement des résidus déhydroalanine et acide déhydroaminobutyrique caractéristiques (**Figure 1.13**). La fragmentation inadéquate du peptide phosphorylé va compliquer son identification et rendre incertaine la position exacte du site de phosphorylation.

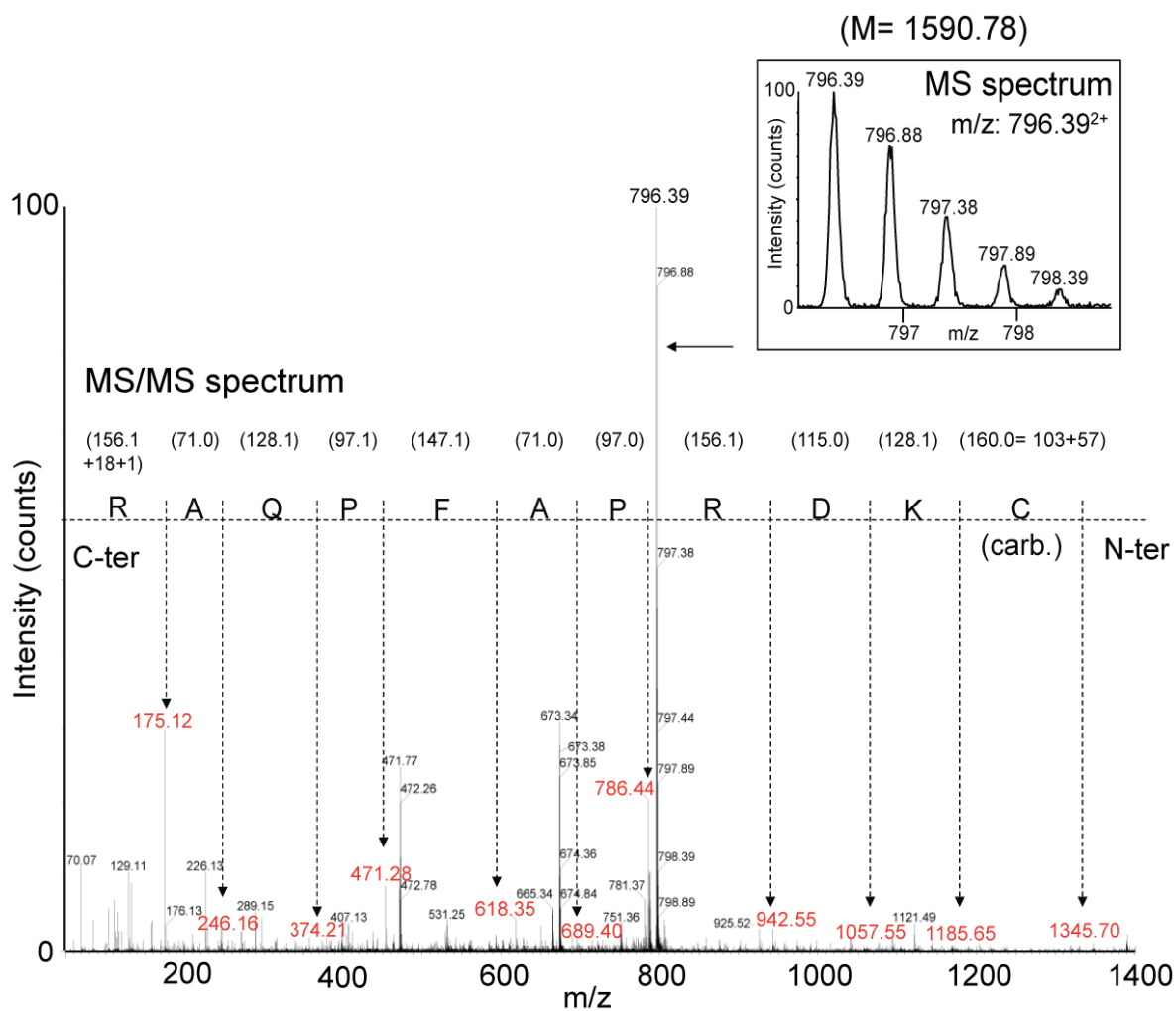


Figure 1.12. Exemple d'un spectre MS/MS en mode CID représentant une série d'ions de type y.

La différence de masse entre chacun des ions fragments correspond à la masse d'un acide aminé spécifique. La séquence sur le spectre est indiquée à partir de l'extrémité C-ter à gauche et montre la portion N-ter à droite ce qui est le schéma conventionnel pour une série d'ions y (indiqués en rouge). (carb: carbamidométhylation).

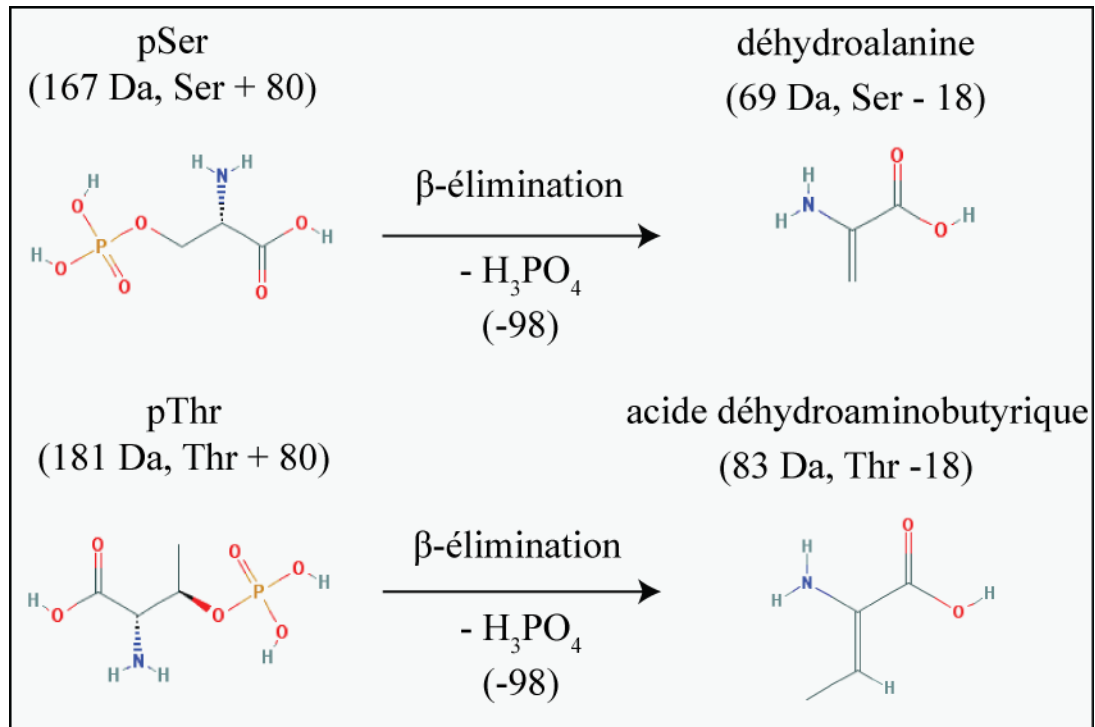


Figure 1.13. Phénomène de β-élimination sur les résidus pSer et pThr suite à la fragmentation CID.

Suite à leur fragmentation en CID, les peptides contenant des pSer et pThr subissent une perte de neutre caractéristique de la phosphorylation. Une β-élimination se produit entraînant la perte de H₃PO₄ (-98 Da). Ainsi, le résidu pSer devient une déhydroalanine et la pThr se transforme en acide déhydroaminobutyrique. Les structures sont obtenues à partir de la base de données PubChem Compound.

Il est possible d'observer des pertes partielles de neutre (80 Da, HPO₃) sur des résidus phosphotyrosines mais le groupement phosphate est plus stable. La phosphorylation des tyrosines est identifiable par la détection spécifique de l'ion immonium (m/z : 216) suite à la fragmentation d'un peptide contenant une phosphotyrosine. Cet ion peut être ciblé plus précisément afin de favoriser son identification[159]. Par ailleurs, la phosphorylation des histidines est très peu souvent identifiée en protéomique étant donné leur nature très labile.

Le groupement phosphate a un effet considérable sur la capacité d'ionisation du phosphopeptide, résultant alors en une diminution de l'intensité du signal MS par rapport

aux peptides non phosphorylés[241]. Afin de réduire cet effet de suppression, il est absolument essentiel de purifier les peptides phosphorylés au sein d'un mélange complexe avant l'injection (**section 1.7.4**). Une attention particulière a été portée afin de trouver de nouvelles stratégies de fragmentation comme la détection de perte de neutre en temps réel qui va déclencher immédiatement une fragmentation supplémentaire (MS^3) afin d'obtenir plus d'information sur la séquence. Cette approche est bénéfique pour les ions les plus abondants. Une méthode alternative est l'utilisation d'une activation à plusieurs niveaux (multi-stage activation) où la fragmentation est induite à la fois sur le précurseur et sur les ions fragments, engendrant alors un spectre hybride regroupant les informations MS^2 et MS^3 qui facilitent l'analyse des données[242].

Les phosphopeptides provenant d'une phosphoprotéine sont identifiés suite à la détection d'une augmentation de masse de 80 Da par groupement phosphate. Il est important de noter que la sulfatation résulte en l'ajout d'une masse très proche (différence de 0.0095 Da). Il est donc essentiel d'avoir des instruments ultra-performants afin d'obtenir une très bonne précision (< 10 ppm) pour les distinguer et éviter les artéfacts dus à la coloration lors d'une approche par gel[148].

Une stratégie intéressante consiste à utiliser un traitement à la phosphatase alcaline afin de cliver de manière spécifique le groupement phosphate du peptide, provoquant une réduction de masse de 80 Da qui peut être détectée en mode MS[243]. Cette stratégie permet d'augmenter la sensibilité de l'analyse et une meilleure fragmentation des peptides déphosphorylés qui sont corrélés par la suite avec leur forme phosphorylée en se basant sur leur m/z et leur temps de rétention.

1.9.3.2. ETD (dissociation par transfert d'électrons)

La fragmentation par ETD (dissociation par transfert d'électrons) est une approche relativement récente complémentaire au CID. Il a été démontré que des anions tels que l'anthracène, l'azobenzène, ou encore le fluoranthène peuvent se transformer en donneur

d'électrons et induire une fragmentation peptidique. Un électron réagit avec un peptide chargé positivement ce qui engendre la neutralisation rapide du site chargé et entraîne la formation d'un radical qui va induire la dissociation de la liaison N-C α [244]. La fragmentation par ETD génère en majorité des ions c et z (voir **Figure 1.11. A**). Elle est de plus en plus utilisée afin d'identifier des modifications post-traductionnelles. ETD est une technique intéressante dans le cas des phosphopeptides étant donné qu'elle ne cause pas la perte de neutre (H₃PO₄) et peut donc augmenter la qualité des spectres MS/MS dans certains cas[244]. En effet, il a été rapporté que la confiance de localisation du site modifié dans la séquence était plus grande[245, 246].

ETD est une méthode optimale pour fragmenter des peptides possédant un haut état de charge (3⁺, 4⁺...), c'est pourquoi il est recommandé de l'utiliser en combinaison avec des protéases qui vont produire des peptides plus longs (Lys-C par exemple). Ainsi, il est plus probable que le peptide possède des charges positives additionnelles dans sa séquence. Le temps de réaction pour ETD est plus long que pour CID, la fragmentation est moins complète et son efficacité dépend aussi du ratio *m/z* du peptide. Pour pallier à ce problème, une méthode par arbre de décision a été implémentée par Swaney, *et al.* afin de guider le mode de fragmentation en temps réel selon le *m/z* du précurseur et son état de charge[247]. Ainsi, il est possible de combiner les avantages de ces deux modes de fragmentation complémentaires en vue d'augmenter le pourcentage de couverture des identifications. C'est la stratégie que nous avons utilisée avec succès dans le **Chapitre 2** afin d'augmenter le nombre de phosphopeptides identifiés[1].

Globalement, ETD est favorable pour la fragmentation d'ions avec des hauts *m/z* et états de charge. Un schéma des règles régissant l'arbre de décision proposé est présenté en **Figure 1.14**.

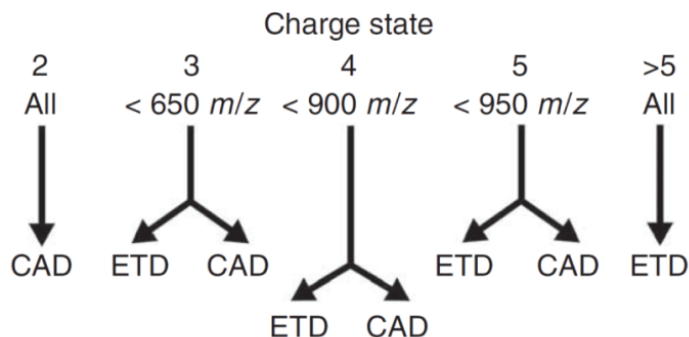


Figure 1.14. Représentation schématique de l'arbre de décision permettant de choisir en temps réel le mode de fragmentation le plus approprié selon le m/z du peptide et son état de charge.

Reproduction de [247] avec la permission de Nature Publishing Group, Nature Methods, copyright 2008.

Il faut par contre garder à l'esprit que le fichier contenant les données résultant de cette analyse doit être divisé selon le mode de fragmentation pour mener les recherches dans les bases de données et minimiser le nombre de faux positifs.

La fragmentation par CID ou ETD génère des ions fragments analysés à basse résolution dans une trappe linéaire tandis que le m/z des précurseurs est obtenue à haute résolution dans l'orbitrap (stratégie « haute-basse »).

1.9.3.3. HCD (dissociation induite par collision de haute énergie)

HCD est la seule méthode à ce jour qui permette d'analyser les ions fragments et leur précurseur en haute résolution (stratégie « haute-haute »)[248]. Elle est accessible avec les instruments de type LTQ-Orbitrap-Velos, Orbitrap Elite et Q Exactive. La fragmentation HCD est obtenue en injectant les ions peptidiques dans une cellule de collision située derrière la trappe en C. Les ions fragments sont ensuite analysés dans l'orbitrap. Cette méthode a déjà fait ses preuves pour l'analyse de phosphopeptides[249] et permet de mieux définir l'état de charge des ions fragments multiplement chargés. Une meilleure confiance

dans la position du site de phosphorylation été aussi rapportée. L'identification de faux positifs est moins probable grâce à une meilleure précision de masse, ce qui permet de diminuer le score du seuil de confiance lors d'analyses statistiques et l'emploi de bases de données inverses. Globalement, bien que le cycle MS-MS/MS soit plus lent en HCD qu'en CID par exemple, le nombre d'identifications finales est égal ou légèrement supérieur[249]. Cette stratégie permet de contourner la règle du « un tiers » qui décrit le fait que les ions dont la masse est inférieure au tiers de la masse du précurseur seront peu détectables dans une trappe linéaire. Ainsi, il est plus facile de détecter les ions immoniums (m/z : 216 Da) générés suite à la fragmentation de phosphopeptides portant une phosphotyrosine.

Il est également possible de conduire des études protéomiques par empreinte peptidique massique (PMF) où seulement les masses des précurseurs sont prises en compte pour leur assignation à une protéine. Dans ce cas, on recherche une correspondance entre les masses expérimentales des peptides tryptiques observés avec les masses théoriques rapportées dans une base de données. Cette stratégie est moins fiable que l'approche en tandem mais permet quand même d'assigner les ions peptidiques caractéristiques d'une protéine connue.

1.9.4. Méthodes de quantification

L'identification de protéines par spectrométrie de masse est devenue une stratégie incontournable durant les dernières années. De plus, cette technique permet une quantification précise ce qui la définit comme la méthode de choix pour réaliser des études dynamiques de phosphorylation.

1.9.4.1. Quantification sans marquage

1.9.4.1.1. Quantification absolue

Il est possible d'obtenir une quantification absolue en incluant des standards internes marqués avec des isotopes lourds dans l'échantillon à analyser. La quantité exacte de chaque standard (marqué au ^{13}C par exemple) est connue et a une masse qui diffère légèrement de celle du peptide endogène tout en ayant les mêmes propriétés physicochimiques. Le standard peut comporter des modifications (phosphorylation, acétylation...) et peut être ajouté avant ou après la digestion enzymatique. Il est analysé par LC-MS simultanément avec l'échantillon afin de prévenir toute variabilité. Cette technologie est rapportée sous le terme AQUA (Absolute Quantification) et a déjà été utilisée avec succès pour quantifier des protéines de faible abondance impliquées dans le silence de gènes chez la levure ou encore pour mesurer précisément un niveau de phosphorylation de la protéine séparase[250]. La **Figure 1.15. A.** illustre la co-élution des espèces endogène et marquée ainsi que leurs profils isotopiques. Les spectres MS/MS des 2 espèces sont représentés en **Figure 1.15. B** et permettent une quantification par SRM (selected reaction monitoring).

Une méthode similaire est l'utilisation de concatamères de peptides protéotypiques QconCAT qui permettent également une quantification absolue. Contrairement à l'approche AQUA, ces concatamères sont synthétisés de manière biologique ce qui laisse une plus grande liberté. Il est ainsi possible d'inclure plusieurs séquences peptidiques différentes dans une seule construction ce qui augmente la robustesse de la quantification[251]. De plus, si les protéines d'intérêt sont connues pour être plus ou moins abondantes, il est possible d'ajuster la stœchiométrie de ces peptides lors de la synthèse en incorporant plusieurs copies d'une séquence spécifique. Le concatamère sera alors introduit dans l'échantillon juste avant la digestion.

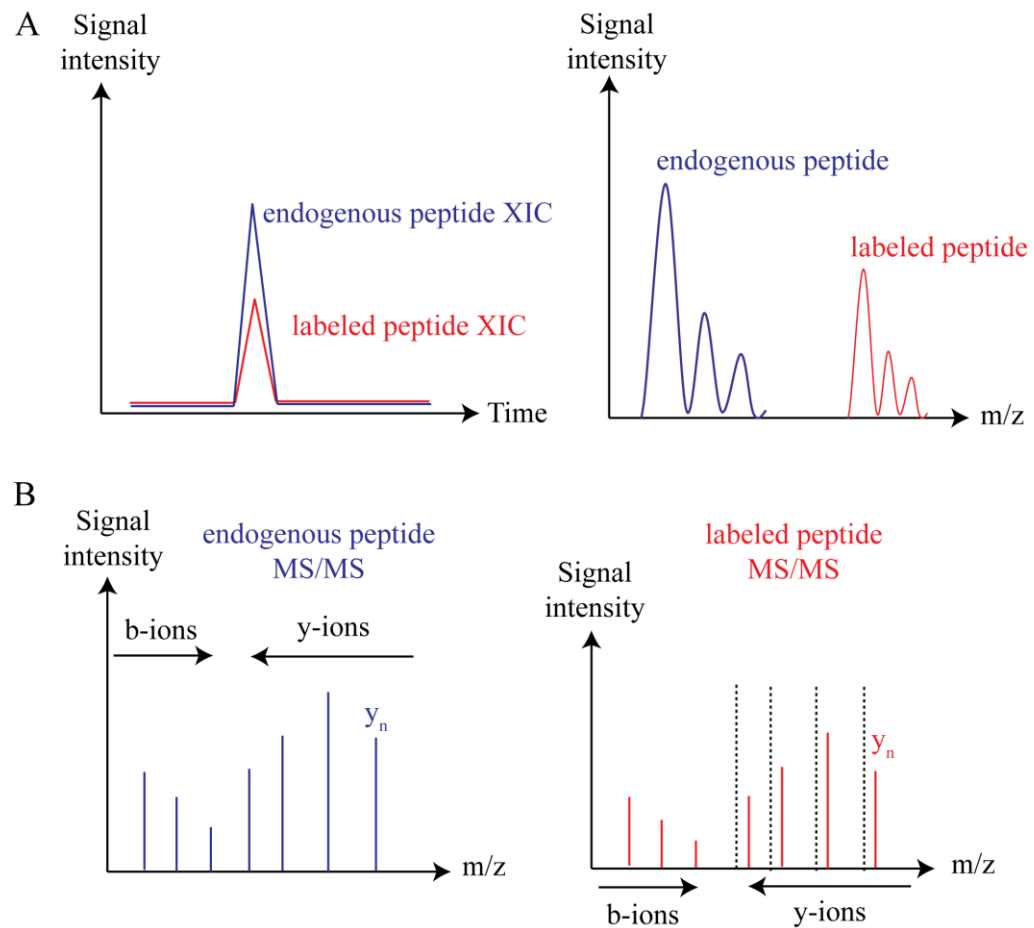


Figure 1.15. Méthodologie de quantification absolue (AQUA) par spectrométrie de masse.

A. Lors d'une approche par quantification absolue en LC-MS, les formes endogènes et synthétiques co-éluent en LC mais sont distinguées par MS grâce à leurs m/z différents. La quantification est réalisée en comparant l'intensité connue du peptide marqué avec celle du peptide endogène. B. Les spectres MS/MS des peptides endogènes et marqués au ^{13}C fragmentent de manière identique mais contiennent certains fragments dont la masse est décalée (l'exemple montre un peptide marqué en C-ter). Il est ainsi possible d'utiliser cette stratégie pour une quantification par SRM (selected reaction monitoring). Adapté de [252] avec la permission de Elsevier, Journal of Proteomics, copyright 2009.

On s'attend à une digestion complète et reproductible étant donné que ces molécules sont très peu repliées et ne possèdent pas de structure tridimensionnelle.

Une quantification absolue au niveau protéique est également possible grâce à la stratégie PSAQ (Protein Standard Absolute Quantification) qui implique des protéines intactes marquées par des isotopes. Ces standards seront ajoutés au début du protocole opératoire afin de contrôler toutes les étapes de préparation de l'échantillon et procure une quantification très précise en évitant par exemple des différences dans la digestion du standard et de l'analyte[253].

D'une manière générale, ces méthodes ne sont utilisées que pour déterminer la quantité de quelques protéines d'intérêt. Un schéma représentatif des méthodes de quantification absolue décrites ci-dessus est présenté en **Figure A1. 4**. Il faut noter que ces méthodes sont assez coûteuses surtout si les synthèses sont réalisées de manière commerciales.

1.9.4.1.2. Quantification relative

Deux méthodes sont généralement utilisées afin d'effectuer une quantification relative sans marquage. La première compare les signaux d'intensité des précurseurs détectés par le spectromètre de masse. Les chromatogrammes sont extraits pour chacun des précurseurs et leurs aires ou leurs intensités sont comparées à travers différentes conditions afin d'obtenir une valeur de changement d'intensité relative[254]. Pour ce type d'analyse, il est essentiel d'effectuer les expériences sur un instrument qui permettent une haute précision de masse et un système de chromatographie très reproductible afin de comparer différentes injections. L'intégration de l'intensité du signal dans le spectre de masse peut se faire de plusieurs façons. On peut utiliser l'intensité ou l'aire du pic monoisotopique (plus simple) ou encore la somme des intensités ou aires des pics isotopiques détectables (plus précise)[255]. Au laboratoire, la méthode de quantification que nous utilisons en routine rapporte l'intensité des peptides et un logiciel (Proteoprofile) a été développé spécialement

afin de quantifier les changements d'intensité observés entre différentes conditions. Il est décrit en détails dans la **section 1.10.3**.

La deuxième méthode, moins précise mais plus directe, consiste à compter et comparer les nombre de spectres MS/MS qui sont acquisitionnés pour chacun des précurseurs, ce qui donne une approximation de leur abondance[256]. Il s'agit par contre de prendre soin de bien ajuster la fenêtre d'exclusion dynamique du spectromètre de masse qui empêche un même précurseur d'être séquencé plusieurs fois dans un intervalle de temps déterminé. Bien qu'intuitive, cette méthode reste controversée car elle repose sur le principe que tous les peptides vont avoir la même réponse linéaire indépendamment de leurs propriétés physiques. Il faut également prendre en compte le phénomène de saturation qui est observé après une certaine limite dans le nombre de spectres acquisitionnés et qui peut être différent d'une protéine à une autre. Ainsi, le rang dynamique pour lequel des changements significatifs peuvent être observés est difficile à déterminer. Des logiciels adaptés tels que Abacus[257] ou Apex[258] permettent de simplifier l'analyse.

Globalement, on estime que la méthode de quantification sans marquage est moins précise que celles avec marquage. Cependant, elle reste la moins coûteuse et est tout de même fiable si l'analyse informatique en aval est réalisée avec une bonne confiance statistique. De plus, il n'y a aucune limite quant au nombre de conditions comparées, le mélange peptidique est moins complexe et le spectromètre de masse ne doit pas fragmenter des peptides redondants avec des marquages différents. En théorie, on peut quantifier des peptides peu ou très abondants de manière plus flexible. C'est la méthode qui est utilisée quotidiennement au laboratoire.

1.9.4.2. **Quantification par marquage métabolique**

Une manière très précise d'obtenir une quantification relative entre deux peptides ou protéines par spectrométrie de masse est d'avoir recours à un marquage isotopique des protéines différent selon les conditions de culture. Ainsi, les protéines provenant des deux

échantillons diffèrent uniquement par la substitution d'un isotope léger par un isotope lourd et se comportent de manière identique. Le ratio d'intensité des deux pics reflète directement les quantités relatives des deux espèces. L'incorporation des isotopes dans les peptides peut se faire de différentes façons. Le marquage peut être réalisé en remplaçant un hydrogène (^1H) par un deutérium (^2H), un ^{12}C par un ^{13}C ou un ^{14}N par un ^{15}N ce qui mène à une différence de masse d'une unité. Peu importe la méthode utilisée, il est important d'atteindre un taux d'incorporation optimal (supérieur à 90 %) pour éviter toute incertitude sur les mesures. Il est en effet primordial de pouvoir comparer des échantillons modifiés avec la même efficacité afin de minimiser les erreurs dues à un marquage incomplet et peu reproductible.

Une méthode qui a fait ses preuves est le SILAC (marquage isotopique stable en culture cellulaire) qui fait intervenir un marquage au niveau d'acides aminés présents dans le milieu de culture pendant la croissance et la division cellulaire[259, 260]. L'expérience est conduite en mettant en culture des cellules contrôles dans un milieu de culture normal et des cellules traitées dans un milieu de culture contenant des $^{13}\text{C}_6$ -arginines ou $^{13}\text{C}_6$ -lysines. L'expérience inverse est également testée pour valider les observations. Le marquage sur les lysines et arginines garantit que la majorité des peptides tryptiques est marquée. Il est également possible de comparer trois conditions simultanément grâce à la technologie SILAC triplex (conditions non-marquée, $^{13}\text{C}_6$, et $^{13}\text{C}_6^{15}\text{N}_4$) ce qui réduit le temps d'analyse mais rend l'échantillon plus complexe. Il faut aussi noter que dans certains modèles cellulaires, la conversion enzymatique de l'arginine lourde en proline est problématique. De plus, cette stratégie n'est pas applicable *in vivo* pour l'étude de patients humains ou dans des modèles animaux par exemple. Une alternative a été trouvée grâce au « Spike-in SILAC » qui utilise un échantillon marqué comme référence[261].

En dehors du marquage au deutérium qui peut causer un léger décalage dans le temps de rétention, les peptides d'intérêt co-éluent en LC et un ratio d'intensité est calculé avec peu d'ambiguïté. L'avantage majeur est que la quantification finale n'est pas

influencée par le protocole opératoire étant donné que les deux échantillons à comparer sont combinés avant la lyse cellulaire.

Une autre approche consiste à introduire un marquage isotopique ^{18}O catalysé par la digestion des protéines avec la trypsine ou la Glu-C[262]. Les isotopes ^{18}O sont incorporés à l'extrémité C-ter des peptides, résultant en une différence de masse de 2 Da par atome. Étant donné que ces enzymes introduisent deux oxygènes, un décalage de 4 Da sera alors visible entre les 2 conditions à comparer. Les artéfacts consécutifs à une réaction chimique sont évités mais il reste néanmoins que le marquage est rarement complet ce qui peut compliquer l'analyse.

1.9.4.3. Quantification par marquage chimique

Les protéines peuvent être marquées après l'étape de culture cellulaire et juste avant la digestion enzymatique. Ceci est réalisé en faisant réagir une molécule chimique avec, par exemple, les groupements thiols ou les amines primaires des peptides. La méthode ICAT (étiquette d'affinité codée par un isotope) permet d'incorporer une étiquette marquée de manière isotopique (8 atomes de deutérium) ou non (8 hydrogènes) liée à un groupement biotine qui réagit avec le groupement thiol des résidus cystéines[263]. Deux marqueurs isotopiques différents sont liés à deux populations protéiques distinctes qui sont ensuite digérées. Seuls les peptides marqués contenant des cystéines sont purifiés grâce au groupement biotine ce qui réduit la complexité de l'échantillon. Cependant, les protéines ne contenant pas de cystéines ne pourront pas être quantifiées.

Le marquage des peptides par iTRAQ (étiquette isobarique pour la quantification relative et absolue de peptide) consiste en l'ajout d'une étiquette isobarique et permet la quantification relative jusqu'à 8 conditions simultanément[264]. La réaction se fait sur les groupements N-ter et les amines des résidus lysines. Les étiquettes sont distinguables uniquement dans le spectre MS/MS suite à la fragmentation du précurseur. C'est une

méthode très pratique lorsqu'on veut analyser une expérience de cinétique par exemple mais la complexité de l'échantillon reste problématique.

La technologie TMT (étiquettes de masse en tandem) fait intervenir le même type de réaction qu'iTRAQ et permet de comparer jusqu'à six conditions en même temps[265]. Chaque réactif comporte un groupement NHS-ester qui réagit avec les amines ainsi qu'un reporteur visible en MS/MS. Un schéma du protocole opératoire est décrit en **Figure A1. 5**.

Ces méthodes conviennent aux échantillons ne pouvant pas être marqués de manière métabolique (organes, biopsies...). Cependant, le marquage par réaction chimique donne lieu à des sous-produits de réaction inattendus qui peuvent influencer la quantification.

Afin de déterminer l'abondance d'une protéine, il est possible de faire la moyenne d'intensité de chacun des peptides détectés. Cependant, dans le cas de l'étude de la phosphorylation, bien souvent on n'identifie qu'un seul peptide portant le site de phosphorylation et on ne peut se fier que sur son intensité afin de le quantifier. Dépendamment des protéines, il faut noter que l'intensité d'un peptide n'est pas exactement proportionnelle à son abondance *in vivo*. En effet, l'accessibilité de la protéase dans la protéine, la solubilité et la capacité d'ionisation de chaque peptide peut être différente et donner lieu à des variations. Cependant, ces changements sont reproductibles, il est donc possible de comparer les intensités des pics entre plusieurs échantillons de manière adéquate. Il faut noter que seul un petit pourcentage de protéines est identifié lors d'une étude protéomique comparativement au protéome entier mais la portion quantifiable est encore plus mince.

La précision de la quantification dépend surtout de la résolution de l'instrument ainsi que du ratio signal sur bruit du pic d'intérêt afin d'être dans des conditions optimales. D'une manière générale, il faut garder à l'esprit que les changements relatifs d'abondance en phosphopeptides peuvent ne pas refléter uniquement la sur- ou sous-phosphorylation d'un site en particulier mais bien le changement global de tous les sites s'il s'agit d'un

peptide multiplement phosphorylé. De plus, les changements globaux d'intensité au niveau de la protéine elle-même peuvent influencer significativement les valeurs observées au niveau du phosphopeptide[266].

1.10. Bioinformatique appliquée aux études protéomiques

Lors d'études protéomiques à grande échelle, des fichiers bruts de plusieurs centaines de mégaoctets sont générés et il est indispensable d'avoir des outils d'analyses performants pour intégrer l'information. Une série de logiciels complémentaires a été développée afin de couvrir tout le processus de traitement depuis l'acquisition des spectres jusqu'à l'interrogation de bases de données. Il est ensuite nécessaire de pouvoir quantifier et interpréter les changements d'intensités peptidiques observés entre plusieurs conditions expérimentales et possiblement y associer une signification biologique.

1.10.1. Traitement des données brutes et interrogation des bases de données

La première étape de traitement bioinformatique consiste à prétraiter les données brutes *.raw* en vue de l'identification des peptides. Les données MS et MS/MS sont rapportées avec un ratio m/z associé à leur intensité détectée par le spectromètre de masse. Simultanément avec l'acquisition des spectres, les données sont transformées sous forme de pic centroïde afin de réduire la taille des fichiers. Les données concernant les précurseurs rapportent uniquement le m/z de la forme monoisotopique. Le bruit de fond est également éliminé et seuls les pics correspondant à des peptides multiplement chargés sont conservés. Les données MS et MS/MS contenues dans le fichier brut sont traitées et transformées dans un format compatible avec l'interrogation des bases de données afin de pouvoir procéder à leur analyse. Au laboratoire, nous utilisons le programme Mascot Daemon afin de convertir les données brutes *.raw* en fichier *.mgf* compatible avec l'engin de recherche Mascot pour

la comparaison avec les bases de données telles qu'Uniprot, NCBI ou IPI. Cette étape est essentielle, par exemple, quand une forme isotopique non majoritaire a été sélectionnée pour la fragmentation. Il est alors nécessaire de faire un ajustement pour faire correspondre le bon m/z avec les ions fragments et minimiser l'identification de faux positifs. Il est également possible d'additionner les intensités de spectres MS/MS similaires afin d'obtenir un meilleur signal et maximiser les chances d'identification. On peut obtenir la séquence des peptides de manière automatisée ainsi que l'identification de la ou des protéine(s) à laquelle elle appartient. Les bases de données contiennent les combinaisons de m/z théoriques des précurseurs et des ions fragments pour toutes les protéines connues chez un organisme déterminé. Elles sont obtenues suite à la digestion théorique de protéines avec diverses enzymes utilisées couramment en protéomique. Une tolérance de masse définie selon l'instrument utilisé est indiquée par l'utilisateur afin de restreindre la recherche. Il est également possible d'inclure des modifications fixes ou variables (oxydation des méthionines, carbamidométhylation des cystéines, déamidation des asparagines et glutamines, phosphorylation des sérines, thréonines et tyrosines...) à prendre en compte lors de l'interrogation. Ce type de modifications peut se produire pendant la préparation de l'échantillon ou *in vivo*. L'utilisateur peut définir une tolérance d'erreur de coupure enzymatique. Dans le cas de Mascot (Matrix Science)[145], l'identification des peptides est rapportée en terme de score de probabilité d'observer par hasard un peptide et ses fragments en fonction de la tolérance de masse et du nombre de fragments détectés. L'algorithme n'est pas publiquement disponible mais correspond à une distribution de probabilité binomiale. Afin de calculer le taux de faux positifs, on soumet les données à une base de données inverse afin de déterminer un score limite acceptable[267]. Cette base de données « non-sens » est constituée des mêmes séquences mais écrites à l'envers. Le taux de faux positifs est donné par le nombre de peptides identifiés au hasard dans la base de données inverse multiplié par deux et divisé par le nombre total d'identifications. On considère en général qu'un taux de faux positifs de 1-2 % est acceptable. Cette approche permet de standardiser les comparaisons entre différentes études protéomiques publiées par divers groupes de recherche[268]. Afin d'attribuer un score d'identification à une protéine,

la manière la plus commune est simplement d'additionner les scores de chacun des peptides qui lui sont assignés. Il faut cependant garder à l'esprit qu'un même peptide peut être partagé par plusieurs protéines.

D'autres algorithmes sont disponibles afin de réaliser ces analyses et se distinguent légèrement par leur système de calcul de pointage au niveau de l'analyse des spectres MS/MS. SEQUEST (Thermo Fisher Scientific)[269] utilise un algorithme de corrélation croisée afin d'attribuer un pointage XCorr qui prend en compte l'intensité et le ratio signal sur bruit des fragments observés puis les compare au modèle théorique. X!Tandem[270] rapporte un score de probabilité « Hyperscore » basé sur l'intensité et le nombre des fragments observés ainsi qu'une valeur d'espérance (E-value) reflétant la probabilité que l'identification soit incorrecte. D'autres outils tels qu'Andromeda implémenté dans le logiciel MaxQuant[271], ProteinProspector ou Spectrum Mill sont aussi disponibles.

Il est aussi possible d'effectuer un séquençage *de novo* en déterminant la séquence d'un peptide sans a priori directement à partir des données brutes.

1.10.2. Outils bioinformatiques dédiés à la phosphoprotéomique

Il est essentiel de pouvoir identifier avec confiance la localisation de la phosphorylation sur un peptide. Cependant, la nature labile du phosphate lors de la fragmentation par CID par exemple complique l'analyse. Si la séquence peptidique contient plusieurs résidus phosphorylables à proximité, il peut s'avérer difficile de distinguer chacun des ions fragments caractéristiques indiquant clairement la position de la modification. De plus, si deux phosphopeptides isomériques co-éluent en LC et sont fragmentés simultanément, il se peut que le spectre MS/MS résulte en une forme mixte qui ne permette pas la localisation du site (**Chapitre 4**). Afin de donner une indication quant à la localisation du site modifié, divers algorithmes ont été développés et attribuent un score reflétant la probabilité de localisation du phosphate. Ascore[272], Phosphoscan[273], MaxQuant[274], Phosphoscore[275] et PhosCalc[276] sont des outils disponibles se basant sur des spectres

MS² (ou MS³ si disponibles) afin de calculer les scores de probabilité et ils ont été conçus initialement pour des fragments obtenus par CID. SLoMo[277], Phosphinator[278], PhosphoRS[279] et Mascot Delta Score[280] ont été développés afin d'être compatibles avec plusieurs modes de fragmentation. Un outil est également intégré dans ProteinProspector et est utilisable pour déterminer la localisation de diverses modifications incluant la phosphorylation[281].

Des bases de données rapportant les sites de phosphorylation connus dans la littérature sont disponibles telles que Phosphosite[282], Phosida[283], PhosphoPep[284], SWISS-PROT[285] et Phospho.ELM[286]. Ces outils facilitent considérablement la recherche de littérature subséquente afin d'associer une fonction à un site précis. Il existe des outils permettant de centrer une séquence peptidique sur le site de phosphorylation et de déterminer un motif environnant. Cela peut indiquer une séquence consensus favorable à la phosphorylation par une kinase plus spécifique ou du moins restreindre les possibilités à une famille particulière. Motif-X[9] et MoDL[287] permettent ainsi d'identifier un motif sur- ou sous-représenté dans un jeu de données ce qui peut constituer un point départ pour l'analyse. Des outils de prédiction de kinases sont également employés couramment afin de déterminer la kinase potentiellement responsable pour la phosphorylation d'un site précis. Ils sont basés sur des algorithmes prenant en compte notamment la séquence consensus environnante du site recherché. On retrouve Prosite[288], NetPhos[289], KinasePhos[290], PPSP[291], PredPhospho[292] et Scansite[293]. Ils ne sont pas reconnus pour être très spécifiques et les résultats des prédictions doivent être considérés avec précaution. NetworKIN[294] utilise par contre des données provenant exclusivement d'interactions kinases-substrats décrites dans la littérature et est beaucoup plus fiable. Des outils permettant de situer le site de phosphorylation dans un domaine particulier sont utiles afin de visualiser si le site d'intérêt se trouve dans une région d'interaction connue ou dans une boucle d'activation. InterPro[295] ou Protein Data Bank (PDB)[296] permettent de visualiser les domaines et PDB répertorie également les structures tridimensionnelles publiées afin de modéliser, par exemple, l'interaction d'un site de phosphorylation avec une structure particulière dans la protéine.

Lorsque les études sont réalisées chez la drosophile, il est important de s'intéresser à la conservation des sites de phosphorylation identifiés. Ceci augmente la probabilité d'avoir une fonction biologique associée pertinente qui aurait été préférentiellement conservée au cours de l'évolution. Il convient donc de réaliser des alignements de séquences entre espèces à l'aide de l'outil blast-p de NCBI ou encore par l'intermédiaire de logiciels proposés par ExPASy. La création de réseaux d'interactions protéiques est une approche intéressante permettant de regrouper les identifications protéiques intervenant dans une même voie de signalisation. Des outils comme String[297] permettent la construction de réseaux d'interactions protéiques incluant des interactions directes (physiques) ou indirectes (fonctionnelles) provenant d'un contexte génomique, d'expériences à grande échelle, de co-expression ou encore de la littérature. L'utilisateur peut définir le niveau de confiance désiré afin de filtrer les données de manière plus ou moins stricte. Cytoscape permet de visualiser et mettre en forme les données selon des regroupements hiérarchiques[298]. L'association de termes GO (Gene Ontology) aux protéines peut indiquer l'effet global d'un stimulus sur une fonction particulière par exemple. Les termes GO sont classifiés en trois catégories qui décrivent l'appartenance des protéines à un processus biologique, un compartiment cellulaire ou une fonction moléculaire. Cependant, il faut noter que chaque protéine est associée à plusieurs termes GO regroupés sous forme ordonnée. De plus, les données reflétant la modulation de sites de phosphorylation ne sont pas toujours représentées de manière significative avec cette stratégie qui est plutôt applicable au niveau protéomique ou génomique. Divers outils sont disponibles afin de mener ces analyses (Panther[299], Amigo[300]...).

Afin de regrouper la plupart des outils décrits ci-dessus en un seul et même logiciel, Mathieu Courcelles a développé au laboratoire Proteoconnections[199] qui a été utilisé tout au long de cette thèse. Proteoconnections est une plateforme d'analyse bioinformatique dédiée à l'interprétation des données issues d'expériences phosphoprotéomique générées par spectrométrie de masse. Elle sert d'abord de base de données et de support afin de stocker les identifications générées lors de différents projets mais permet aussi d'évaluer la qualité des analyses. Elle intègre une multitude d'outils bioinformatiques pour faciliter

l'exploration du phosphoprotéome et proposer de nouvelles hypothèses biologiques fonctionnelles.

1.10.3. Logiciel Proteoprofile pour la quantification sans marquage

Proteoprofile est un outil qui regroupe 4 modules afin d'analyser et quantifier les informations provenant des données brutes générées par LC-MS/MS. Il permet de corréler l'abondance des peptides à travers les triplicatas, conditions et fractions en regroupant les données d'identification provenant des bases de données avec les données d'intensité. L'utilité de ce logiciel réside en sa capacité de regrouper des peptides identiques provenant de plusieurs injections mais qui n'ont pas été forcément séquencés ou assignés à une protéine dans chaque analyse. Ainsi, il est possible d'obtenir des données quantitatives plus complètes et représentatives. Le premier module a pour but la conversion des données brutes *.raw* en carte peptidique *.sdf* représentant le *m/z* du peptide sur l'axe des x, le temps de rétention sur l'axe des y et un code de couleur rapporte son intensité sur l'axe des z. L'intensité du pic monoisotopique est rapportée pour la quantification. Un fichier *.csv* est généré et regroupe les abondances de chaque ion corrélées avec leur *m/z* et temps de rétention. Le second module permet de convertir le fichier *.csv* provenant de Mascot qui contient les identifications peptidiques en une liste d'identifications non redondantes. Le troisième module va regrouper les données d'intensités provenant des cartes avec les identifications. Après l'alignement des cartes selon leurs coordonnées *m/z* et temps de rétention, il est possible de corréler les données d'intensités détectées dans chaque condition en précisant une certaine tolérance de masse et de décalage de temps de rétention. Il est possible d'appliquer une étape de normalisation entre les triplicatas au besoin. Dans le cas d'une expérience en 2D-LC, l'intensité de chaque ion détecté dans chacune des fractions est additionnée pour calculer l'intensité globale. Dans le cas d'une expérience par FAIMS, uniquement le CV permettant la plus grande transmission de l'ion sera prise en considération pour l'intensité finale. Finalement, la quatrième étape permet à l'utilisateur de visualiser et valider les données d'abondance en inspectant chaque peptide

individuellement. Ceci constitue un aspect unique à Proteoprofile. Ainsi, les erreurs éventuelles peuvent être corrigées afin de calculer des changements d'abondances significatifs et fiables. Un schéma du processus est décrit en **Figure 1.16**. D'autres logiciels similaires à Proteoprofile sont également disponibles commercialement tels que QuantLynx (Waters) ou SIEVE (Thermo Scientific). D'autres systèmes ont été développés au sein d'autres laboratoires tels que MaxQuant[271], MassUntangler[301], RIBAR[302], Census[303] ou encore ProtQuant[304].

Lors d'études comparatives, la moyenne d'intensité entre chacun des triplicatas de chacune des conditions est déterminée et le ratio d'intensité moyenne entre deux conditions est calculé afin de déterminer un changement d'expression d'un peptide. Il convient de donner une valeur statistique à la variation observée et nous utilisons le test T de Student pour définir notre confiance (l'hypothèse nulle est $\bar{X} = \bar{Y}$).

La statistique T est donnée par :

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}}$$

Équation 3.

où \bar{X} et \bar{Y} représente les moyennes des 2 conditions, S la variance et n le nombre de réplicatas.

La distribution T de Student permet de déterminer la probabilité d'observer la valeur T obtenue et de calculer la valeur p (p-value), soit la probabilité d'obtenir la même valeur ou une valeur plus extrême. La valeur p est le plus petit niveau auquel on rejette l'hypothèse nulle, soit la probabilité d'obtenir un faux positif.

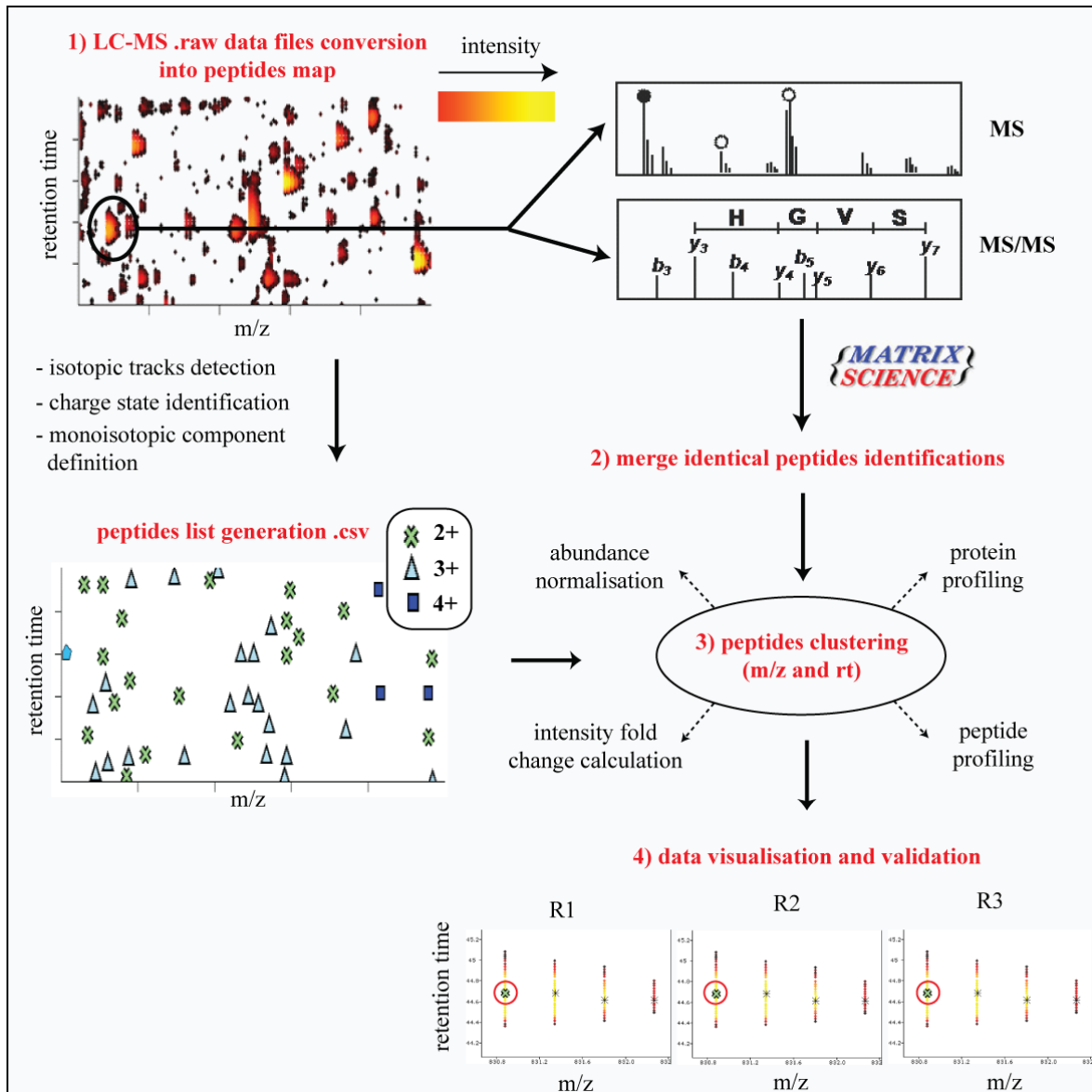


Figure 1.16. Représentation des 4 modules du logiciel Proteoprofile utilisé lors de la quantification de données protéomiques sans marquage.

1) Les données brutes *.raw* sont converties en carte peptidique et une liste des abondances de chaque ion est générée en corrélation avec son *m/z* et temps de rétention. 2) Les identifications obtenues à partir de Mascot sont filtrées de manière non redondante et triées pour ne garder que les entrées uniques. 3) Les données d'intensité et d'identification sont regroupées dans un même fichier afin de déterminer des changements d'abondances peptidiques ou protéiques entre différentes conditions. 4) Les données sont visualisées et validées afin de corriger toute erreur éventuelle de détection.

D'une manière générale, on considère qu'une valeur p inférieure à 0.05 est acceptable (soit 5 % de chance d'obtenir un faux positif).

1.11. Objectifs de recherche

L'objectif principal de cette thèse est d'identifier les événements de phosphorylation intervenant dans la voie de signalisation de l'insuline. Pour y parvenir, nous avons mené deux études phosphoprotéomiques globales grâce à une approche combinant FAIMS et LC-MS/MS.

Lors d'études phosphoprotéomiques par spectrométrie de masse, il convient de réduire la complexité de l'échantillon avant son analyse afin de maximiser les chances d'identification par LC-MS/MS. Communément, on utilise la chromatographie liquide en deux dimensions (SCX/C₁₈) afin de séparer les phosphopeptides avant leur injection. Cependant, bien que l'instrumentation soit de plus en plus performante, seulement 20 à 30 % des peptides présents seront fragmentés. Le problème à résoudre consiste à augmenter le nombre de phosphopeptides identifiés par spectrométrie de masse en employant une méthode de séparation orthogonale à la LC-MS : le FAIMS. Cette approche diminue la complexité de l'échantillon tout en réduisant le bruit de fond ce qui permet d'augmenter le ratio signal sur bruit et la gamme dynamique d'identification.

La première analyse a pour but de démontrer les bénéfices de l'utilisation du FAIMS comparativement à l'approche par électronébulisation conventionnelle lors d'études phosphoprotéomiques à grande échelle. Nous mettons à profit cette technologie afin de disséquer les mécanismes moléculaires intervenant dans la voie de l'insuline chez la drosophile. Cette voie fait intervenir des cascades de phosphorylation suite à l'activation d'un récepteur tyrosine kinase (RTK). Il est bien connu que plusieurs RTKs sont activés anormalement dans de nombreux types de cancers et nous avons choisi cette approche dans le but de mieux comprendre les événements de phosphorylation impliqués en aval. Bien que certaines phosphoprotéines soient déjà caractérisées au sein de cette voie, ce projet est

mené d'une manière exploratoire afin d'obtenir une vue globale de la voie. L'accomplissement de ce premier objectif nous a encouragés à étudier plus en profondeur les impacts directs de l'insuline sur certaines kinases telles que la protéine kinase CK2.

Ainsi, la deuxième étude a pour but la combinaison d'une approche RNAi avec la spectrométrie de masse afin de mieux comprendre le rôle de CK2 dans la voie de l'insuline. Le problème à résoudre consiste à comprendre pourquoi près de la moitié des phosphoprotéines modulées par la stimulation à l'insuline comporte un motif consensus CK2. Notre objectif est d'étudier le phosphoprotéome différentiel des cellules S2 contrôles et stimulées à l'insuline avec ou sans traitement par RNAi visant l'inhibition de CK2. Cette analyse a pour but d'identifier les protéines dont la phosphorylation est modulée par l'insuline et par CK2 (de manière directe ou indirecte). Ensuite, un de nos objectifs est de confirmer les cibles directes de CK2 grâce à des essais kinase *in vitro* afin d'appuyer les résultats observés dans les études globales. Finalement, la dernière hypothèse à vérifier consiste à mesurer l'activité de CK2 en réponse au traitement à l'insuline.

Suite à la réalisation des deux objectifs précédents, un troisième sujet a été abordé dans cette thèse et consiste en l'étude de phosphopeptides isomériques (phosphopeptides ayant la même séquence mais dont la position du phosphate varie). L'analyse des résultats générés lors des études phosphoprotéomiques chez la drosophile nous a menés à nous intéresser à la présence de ces phosphoisomères pour lesquels peu d'information est disponible dans la littérature (fréquence, méthodes de séparation...). Cet aspect mérite d'être documenté car la localisation précise du site de phosphorylation est d'une importance primordiale afin de pouvoir lui attribuer une fonction biologique et envisager des expériences de validation fonctionnelle. La problématique à résoudre consiste à déterminer la proportion de phosphoisomères à l'aide de deux algorithmes permettant de détecter les isomères séparés par LC ainsi que ceux qui co-éluent. Il s'agit ensuite de cibler les isomères potentiels grâce à une liste d'inclusion lors d'une seconde injection afin de confirmer par MS/MS la position du site modifié.

Chapitre 2 : Improvement of phosphoproteome analyses using FAIMS and decision tree fragmentation; Application to the insulin signalling pathway in *Drosophila melanogaster* S2 cells

Gaëlle Bridon^{1,2,5}, **Éric Bonneil**^{1,5}, Tara Muratore-Schroeder^{1,4},
Olivier Caron-Lizotte¹, Pierre Thibault^{1,2,3}

Article publié dans *Journal of Proteome Research*, 2012

J Proteome Res. 2012; 11(2): 927-40.

[1]

Inclus avec la permission de John Wiley and Sons

¹Institute for Research in Immunology and Cancer, ²Department of Biochemistry, Université de Montréal, P.O. Box 6128, Station. Centre-ville, Montréal, Québec, Canada H3C 3J7, ³Department of Chemistry, Université de Montréal, ⁴Present address: Thermo Fisher Scientific, 265 Davidson Ave, Suite 101 Somerset, NJ 08873.

⁵These authors have contributed equally to this work.

2.1. Contribution des auteurs

Gaëlle Bridon, Éric Bonneil et Pierre Thibault ont écrit le manuscrit. **Gaëlle Bridon** a effectué la culture cellulaire, la préparation des échantillons, l'analyse par spectrométrie de masse ainsi que le traitement des données. Éric Bonneil et Tara Muratore-Schroeder ont participé à la génération des données comparatives et Éric Bonneil a implémenté l'instrumentation. Olivier Caron-Lizotte a programmé le logiciel Proteoprofile servant à l'analyse bioinformatique des données. Pierre Thibault a supervisé cette analyse.

2.2. Abstract

This report examines the analytical benefits of high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to liquid chromatography mass spectrometry (LC-MS) for phosphoproteomics analyses. The ability of FAIMS to separate multiply-charged peptide ions from chemical interferences confers a unique advantage in phosphoproteomics by enhancing the detection of low abundance phosphopeptides. LC-FAIMS-MS experiments performed on TiO₂-enriched tryptic digests from *Drosophila melanogaster* provided a 50% increase in phosphopeptide identification compared to conventional LC-MS analysis. Also, FAIMS can be used to select different population of multiply-charged phosphopeptide ions prior to their activation with either collision activated dissociation (CAD) or electron transfer dissociation (ETD). Importantly, FAIMS enabled the resolution of co-eluting phosphoisomers of different abundances to facilitate their unambiguous identification using conventional database search engines. The benefits of FAIMS in large-scale phosphoproteomics of *D. melanogaster* are further investigated using label-free quantitation to identify differentially regulated phosphoproteins in response to insulin stimulation.

2.3. Introduction

Protein phosphorylation is a highly dynamic and reversible modification that confers structural changes on their substrates and influences their cellular outcome via protein complex formation, translocation, activation or inactivation. This post-translational modification (PTM) also plays an important role in a wide range of cellular processes including cell division, growth and differentiation. In eukaryotes, phosphorylation is mostly found on hydroxylated amino acids such as serine, threonine, and tyrosine residues (*O*-phosphorylation), though this modification has been reported to take place to a much lower extent on basic residues (*N*-phosphorylation)[305]. Protein phosphorylation is finely orchestrated by a network of protein kinases and phosphatases that add or remove a phosphate group from specific substrates. Misregulation of these signalling network by mutation or abnormal protein expression is the cause of numerous diseases including cancer[306].

The large-scale profiling of protein phosphorylation and the identification of modification sites represent challenging tasks in view of the relatively low abundance of this modification in protein digests. Remarkable analytical advances have been realized over the past decades to facilitate the identification of protein phosphorylation in large-scale proteomics experiments. One of the most important developments came from the availability of affinity media for the specific enrichment of low abundance phosphopeptides from complex cell extracts. Large-scale phosphoproteomic studies often use immobilized metal affinity chromatography (IMAC)[307-310] or metal oxide (e.g. TiO_2)[142, 311, 312] media to selectively purify phosphopeptides. IMAC and metal oxide media offer complementary selectivity for phosphopeptides enrichment, and previous report described their sequential use to enhance the comprehensiveness of phosphoproteome analyses[171]. These significant advances have greatly contributed to the expanding repertoire of protein phosphorylation deposited in public repositories that currently counts more than 25 000 sites for the human proteome alone[313]. However, these numbers appear relatively modest compared to the 100 000 sites estimated for the human phosphoproteome[4], and

underscores several issues relating to the comprehensiveness and dynamic range of current identification platforms.

Several analytical approaches have been developed to accommodate the overwhelming biological complexity commonly found in phosphopeptides extracts. Improvement of separation capacity has been achieved using multi-dimensional chromatography, with strong cation exchange (SCX) chromatography being the most popular first dimension for large-scale phosphoproteomic experiments. This has been successfully implemented either before[314, 315] or after[142] TiO₂ enrichment. The enhanced sample loading and peak capacity available with the combined SCX/C₁₈ chromatography typically provide more than 5-fold enhancement in phosphopeptide identification compared to LC-MS systems using reverse phase chromatography alone. Improvement in phosphopeptide identification has also been achieved using collision activated dissociation (CAD) and electron transfer dissociation (ETD) whereby the selection of the activation mode is tailored according to the charge state and the m/z value of precursor ions. ETD and CAD are highly complementary, with only 18% of peptides being common to both activation modes[316].

While phosphoproteome coverage has been greatly expanded using multi-dimensional chromatography and combined activation modes, the sequencing capability of modern MS instruments is often limited by the intensity of the precursor ion and the presence of confounding ions of similar m/z values that give rise to mixed product ion spectra. Numerous types of interfering ions such as column bleed, solvent clusters, co-eluting peptide ions, or phosphoisomers can compromise the quality of the corresponding MS/MS spectra. In this context, ion mobility spectrometry provides a complementary approach by separating gas-phase ions based on their mobility in electric fields. Previous investigations including work from our group have documented the use of ion mobility selection (IMS) to improve the quality of MS/MS spectra, ultimately leading to higher identification rate and sequence coverage in large-scale proteomics[230, 231, 233, 234, 317-319]. High field asymmetric waveform ion mobility spectrometry (FAIMS) is a type of

IMS that is based on the difference of mobility at high and low electric fields. In FAIMS, ions are entrained by a carrier gas (e.g. a mixture of nitrogen: helium) in a gap between two electrodes to which is applied a high voltage asymmetric waveform. Ion selection is achieved by applying a compensation voltage (CV) that is superimposed to the waveform in order to transmit specific ions that would otherwise be neutralized on an electrode. Ions of different mobility at high vs. low fields are transmitted in turn by scanning the CV voltage. In the case of peptide analysis, FAIMS has emerged as separation technique orthogonal to MS, and enables the resolution of isomers and positional variants [216, 223, 320]. In combination with LC-MS, FAIMS can be used to enhance the transmission of multiply-charged ions, and to improve the signal to noise ratio by an order of magnitude[233, 234]. Although the resolution of commercial FAIMS unit is relatively modest ($R_s \sim 15$), it provides partial separation of phosphoisomers and favours the unambiguous identification of modification sites using ETD[321, 322].

In the present work, we evaluated the analytical merits of FAIMS device coupled to a LTQ-Orbitrap for phosphoproteomics analyses of *Drosophila melanogaster* S2 cells. We compared the overall gain in phosphopeptide identification between conventional LC-MS and LC-FAIMS-MS with sequential CV stepping. We also evaluated the analytical benefits of FAIMS for the separation of phosphoisomers that remained unresolved under conventional LC-MS conditions. Applications of FAIMS to quantitative phosphoproteomics are further demonstrated for *D. melanogaster* S2 cells to probe the rapid changes of protein phosphorylation following cell stimulation by insulin.

2.4. Materials and methods

2.4.1. Materials

Modified porcine sequencing grade modified trypsin was obtained from Promega (Madison, WI, USA). Acetonitrile (ACN) and HPLC grade water was purchased from Fisher Scientific (Whitby, ON, Canada). Ammonium bicarbonate and formic acid were obtained from EM Science (Mississauga, ON, Canada). Ammonium hydroxide, trifluoroacetic acid (TFA), DTT (DL-dithiothreitol), iodoacetamide, sucrose, MgCl₂, CaCl₂, HEPES, NaCl, glycerol, protease inhibitor cocktail (4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), pepstatinA, E-64, bestatin, leupeptin, and aprotinin), phosphatase inhibitor cocktail (sodium vanadate, sodium molybdate, sodium tartrate, and imidazole), serum-free and protein-free insect medium-1 and human insulin solution were purchased from Sigma-Aldrich (Oakville, ON, Canada). Bradford protein assay and Criterion XT precast gels (4-12% Bis-Tris) were obtained from Bio-Rad (Mississauga, ON, Canada). Lactic acid was purchased from Fluka (Saint Louis, MO). Tris was purchased from EMD Omnipur (Lawrence, KS). Titanium dioxide (Titanium Bulk, 5 µm, 500 mg) was obtained from Canadian Life Science (Peterborough, ON). Bond breaker TCEP (Tris[2-carboxyethyl] phosphine) was purchased from Pierce Biotechnology Inc. (Rockford, IL). Phosphate buffered saline (PBS) and ethylene-diamine-tetra-acetic acid (EDTA) were obtained from HyClone (Thermo Scientific, Logan, UT). Oasis HLB cartridges (1cc, 30 mg) were purchased from Waters (Milford, MA). ECL chemiluminescence detection system was purchased from Amersham Pharmacia Biotech (Montréal, QC, Canada). Actin, p-ERK (1/2) (T₂₀₂/Y₂₀₄) and p-AKT (S₄₇₃) primary antibodies were purchased from Cell Signaling (Danvers, MA). Goat anti-mouse IgG and goat anti-rabbit IgG secondary antibody were purchased from Millipore (Ottawa, ON, Canada). Capillary columns for LC-MS were packed in-house using Jupiter C₁₈ (3 µm) particles from Phenomenex (Torrance, CA), and fused silica tubing from Polymicro Technologies (Phoenix, AZ).

2.4.2. Cell culture and protein extraction

Drosophila melanogaster Schneider S2 cells were cultured in serum-free and protein-free insect medium at 28 °C. Cells were plated in 100-mm Petri dishes at a density of 18 million cells/dish for 24 hours before treatment. For insulin stimulation, cells were treated from 4 min up to 60 min with a final concentration of insulin ranging from 0.43 to 3.44 μ M. Total cell lysates were obtained using a detergent-free cell lysis. Cells were washed with PBS and then cell lysis was performed in an extraction buffer (20 mM Hepes, 1.5 mM MgCl₂, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol) using a dounce metal homogenizer to extract total proteins. After centrifugation at 13 000 g, the supernatant was isolated and the membrane pellet was discarded. During buffer preparation, 1mM DTT, proteases and phosphatases inhibitor cocktails were added. Total protein amount was quantitated using Bradford protein assay.

2.4.3. Immunoblotting analysis

Lysates were boiled for 5 min in Laemmli buffer (10% (w/v) glycerol, 2% SDS, 10% (v/v) 2-mercaptoethanol and 0.0625 M Tris-HCl, pH 6.8), and separated by SDS-PAGE prior to electroblotting onto PVDF membrane. After blocking of non-specific binding sites with 5% non-fat milk, membranes were incubated with primary antibody (p-ERK(1/2), p-AKT or Actin) followed by horseradish peroxidase-conjugated secondary antibody. The secondary antibody was detected with the ECL chemiluminescence detection system.

2.4.4. Trypsin digestion

Proteins were reduced in 50 mM ammonium bicarbonate containing 0.5 mM TCEP and 1% SDS for 20 min at 37°C, then alkylated in 5 mM iodoacetamide for 20 min at 37°C. Excess iodoacetamide was neutralized using 5 mM DTT. After dilution with 50 mM ammonium

bicarbonate, proteins were digested overnight in 0.1% SDS with sequencing grade modified trypsin (enzyme/protein ratio: 1/50) at 37 °C with high agitation. The digest mixture was acidified with TFA, and then evaporated to dryness in a SpeedVac.

2.4.5. Phosphopeptide isolation

For the TiO₂ enrichment procedure, sample loading, washing and elution were performed by spinning the micro-column at 8 000 rpm at 4°C in a regular Eppendorf microcentrifuge. The spinning time and speed were adjusted as a function of the elution rate. The 2,5-dihydroxybenzoic acid (DHB) usually used to resuspend the sample was replaced by 250 mM lactic acid to avoid interference in 2D-LC analysis. For the 2D-LC-MS analysis, 6 mg of titanium dioxide phase was packed in a tip and 1.5 mg of protein digest was enriched on two micro-columns (750 µg/micro-column). After drying, the digest mixture was dissolved in 400 µL of 250 mM lactic acid (3% TFA/70% ACN), centrifuged for 5 min at 13 000 rpm and the soluble supernatant was loaded on 2 micro-columns previously equilibrated with 100 µL of 3% TFA/70% ACN. Each micro-column was washed with 100 µL of lactic acid solution followed by 200 µL of 3% TFA/70% ACN to remove non-specific binding peptides. Phosphopeptides were eluted with 200 µL of 1% NH₄OH pH 10 in water and acidified with 7 µL of TFA. Eluates from two TiO₂ micro-columns were combined and desalted using Oasis HLB cartridges by spinning at 1 200 rpm at 4°C. After conditioning with 1 mL of 100% ACN/0.1% TFA and washing with 0.1% TFA in water, the sample was loaded, washed with 0.1% TFA in water, then eluted with 1 mL of 70% ACN (0.1% TFA) prior to evaporation on a SpeedVac. The equivalent of 1.5 mg of starting material was resuspended in 16 µL of 0.2% formic acid (2% ACN) for 2D-LC-MS/MS experiments on the LTQ-Orbitrap mass spectrometer. Samples were prepared in biological triplicate for each condition, and one third of each replicate was mixed together to obtain higher proteome coverage in 2D-LC-MS/MS experiments.

2.4.6. Liquid chromatography

For LC experiments, protein digests were separated using a home-made 10 cm x 150 μm i.d. analytical column and a 4 mm length, 360 μm i.d. trap column packed with 3 μm C₁₈ particles (Jupiter 300 Å, Phenomenex, Torrance, CA). The 2D-LC-MS experiments comprised a strong cation exchange (SCX) column (1 mm, Optiguard, Optimize Technologies, Oregon City, OR) on line with the C₁₈ trap column. Peptides were eluted from the SCX column using six different ammonium acetate salt fractions at pH 3.0 (0, 250, 500, 750, 1 000, 2 000 mM) in 0.2% formic acid (2% ACN). Tryptic digests were injected on the column and elution was achieved using a linear gradient of 5-40% acetonitrile (0.2% formic acid) in 51 min with a flow rate of 600 nL/min for both FAIMS and non-FAIMS experiments.

2.4.7. FAIMS interface and mass spectrometry

The FAIMS interface was coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) via a nanoelectrospray source. The FAIMS electrodes were operated at a dispersion voltage of -5 000 V with a 3.5 L/min flow of helium/nitrogen 1/1. The temperature of the inner and outer electrodes was set to 70 and 90°C, respectively. The Ion Max source of the LTQ-Orbitrap was modified for nanoelectrospray ionization. For infusion experiments, samples were introduced at a 600 nL/min flow rate using the syringe pump on the LTQ-Orbitrap XL instrument. Optimal transmission of multiply charged peptide precursor ions was determined by scanning the CV from -50 V to -10 V at DV - 5000 V, while the sample was infused. For CV stepping in LC-FAIMS-MS experiments, the Xcalibur software enables a programmable CV function to be synchronized with the operation of the LTQ-Orbitrap XL. CV stepping experiments consists of cycling through 5 CV values at a rate of 1s/CV step with 0.3 s interscan time while the corresponding data for each CV is acquired by the mass spectrometer. All MS analyses were performed using an LTQ-Orbitrap XL hybrid mass spectrometer with a nanoelectrospray ion source coupled to

a LC system (Eksigent, Dublin, CA) equipped with a Finnigan AS autosampler (Thermo Fisher Scientific). The conventional MS spectra (survey scan) were acquired at high resolution ($M/\Delta M$: 60 000 full width half maximum) over the acquisition range of m/z : 400-1600. Unless otherwise indicated, MS/MS spectra were collected using data-dependent acquisition for multiply charged ions exceeding a threshold of 10 000 counts. The ETD reagent target value was set to 100 000 and the reaction time was 50 ms.

2.4.8. Database searching and phosphoproteomics analyses

MS data were analyzed using the Xcalibur software (version 2.0.7). Peak lists were then generated using the Mascot distiller software (version 2.3.2.0, Matrix science) where MS processing was achieved using the LCQ_plus_zoom script. Database searches were performed using the search engine Mascot (version 2.2.0, Matrix Science, London, UK) and used the Uniprot *Drosophila melanogaster* (Dmel) forward/reverse database containing 58 904 entries (version 3.54, released November 2006). The error window for experimental peptide mass values and MS^2 fragment ion mass values were set to +/- 15 ppm and 0.5 Da, respectively. The number of allowed missed cleavage sites for trypsin was set to 1 and phosphorylation (STY), oxidation (M), carbamidomethyl (C), and deamidation (NQ) were all selected as variable modifications. No fixed modification was included in the search. A cut-off score threshold was established for a false-discovery rate of less than 2% ($p < 0.02$)[323]. All peptide identifications were transferred to ProteoConnections, a bioinformatics platform facilitating the organization and interpretation of proteomics data[324]. The confidence in the location of phosphorylation sites is determined using the probability of a specific assignment based on the method described by Olsen, *et al.*[19]. Phosphorylation sites were also annotated according to consensus motifs and kinase prediction tools (e.g. PPSP, NetworKIN, etc...) integrated within ProteoConnections. Over-represented patterns from phosphopeptide sequences were extracted using Motif-X software (<http://motif-x.med.harvard.edu/motif-x.html>, version 1.2 10.05.06, 2005)[9].

Identified phosphosites were centered with six flanking amino acids on each side of the modified residue. Parameters set for the width, occurrences, significance and background were 13, 20, 0.000001 and Flybase + Drosophila Proteome, respectively.

2.4.9. Peptide detection, clustering and validation

The comparison of peptide abundances across different experimental paradigms was achieved using label free quantitative proteomics[233, 325]. Briefly, raw data files (*.raw*) from the Xcalibur software were converted into peptide map files representing all ions according to their corresponding m/z values, retention time, intensity and charge state. Intensity values above a threshold of typically 10 000 counts were considered for further analysis. Peptides abundances were assessed using the peak top intensity values. Intensities of peptides eluting across several SCX fractions were summed together. For FAIMS experiments, only the CV allowing the maximal ion transmission was considered to calculate peptide intensity. Clustering of peptide maps across different sample sets were performed on peptides associated to a Mascot entry using hierarchical clustering with tolerances of ± 15 ppm and ± 1 min for peptide mass and retention time, respectively. Normalization of retention time is then performed on the initial peptide cluster list using a dynamic and nonlinear correction that confines the retention time distribution to less than 0.1 min ($< 0.3\%$ RSD) on average. Reproducible changes in abundance across conditions were determined using a two-tail homoscedastic t-test on sample replicates to identify peptide clusters with p-values < 0.1 with fold change greater than 7 standard deviations. Peptide clusters fulfilling these selection criteria were inspected manually to validate identification and changes in abundance. Global intensities were normalized based on the average intensity of all peptides signals.

2.5. Results and Discussion

One of the main challenge in any large-scale proteomics experiment is the overwhelming number of co-eluting ions that often exceeds the MS/MS capacity and the sensitivity of MS instruments. The acquisition constraints under which MS instrument currently operate often result in the under representation of low abundance peptides. While repeated LC-MS/MS analyses or the implementation of iterative exclusion[326] can facilitate the identification of low abundance ions, MS/MS sampling is ultimately limited by the ability to distinguish genuine peptide ions against a confounding background of ionized molecules arising from column bleed, solvent clusters and other low abundance peptides. This difficulty is further exacerbated by the dynamic range of certain type of MS instruments, such as ion traps, that limit the maximum number of ions analyzed simultaneously to maintain analytical performances. Previous reports using FAIMS on a Q-TOF[234] and a LTQ-Orbitrap[233] have shown that mobility selection offers a convenient approach to enhance the transmission of multiply charged ions for proteomics applications. FAIMS not only provides higher signal to noise ratio for peptide detection, but also separate ions that would normally be confounded in conventional MS analysis. We surmised that the inherent analytical benefits of FAIMS for peptide identification would be of immediate benefits in phosphoproteomics to enhance the confidence in site localization and the comprehensiveness of these analyses. Accordingly, we investigated the capability of FAIMS to facilitate phosphoproteomics analyses of *D. melanogaster* using the LTQ-Orbitrap XL mass spectrometer.

2.5.1. Enhancement of phosphopeptide detection in LC-MS/MS experiments using FAIMS

To investigate the analytical potentials of FAIMS in large-scale phosphoproteomics, we first compared the number of phosphopeptides from total cell extracts of *D. melanogaster*

with and without FAIMS using identical chromatographic conditions (**Figure 2.1**). Protein extracts from drosophila S2 cells were lysed, digested with trypsin, and phosphopeptides were enriched with custom TiO₂ columns[142]. Phosphopeptide extracts corresponding to an equivalent of 1.5 mg of starting material were analysed in triplicate using 2D-LC-MS/MS with and without FAIMS. To maximize the transmission of multiply-charged ions across the CV range, we stepped through 5 CV values (i.e. -42 V, -38 V, -32 V, -28 V and -24 V) for a maximum of 5.4 s per acquisition cycle[233]. Clustering analysis performed across fractions and replicates indicated that the number of detected peptides for 2D-LC-MS/MS with and without FAIMS were 44 760 and 37 755, respectively. These analyses revealed that 2D-LC-MS/MS analyses of phosphopeptide extracts using FAIMS provided a 19% increase in the number of detected peptides compared to conventional 2D-LC-MS/MS experiments. A comparison of the reproducibility of peptide intensities indicated that 2D-LC-MS/MS with and without FAIMS provided RSD values of 45% and 42%, respectively. A plot showing the distribution of peptide intensities in each experiment is presented in **Figure A2. 1**, and attest of the reproducibility of the label-free quantitation for the analysis of these sample replicates. These results are consistent with previous studies, where an increase in the number of multiply charged ions was observed for lower abundance ions (typically less than 50 000 counts)[233, 234].

To evaluate the impact of FAIMS on phosphopeptide identification, we compared the results obtained from 2D-LC-MS/MS experiments performed with and without FAIMS using a data-dependent decision tree algorithm that exploit the complementary ion activation modes of CAD and ETD[316]. We identified a total of 6 122 unique peptides in all FAIMS and non-FAIMS experiments, of which 4 445 were phosphopeptides (enrichment level of 73%). For convenience, all identifications can be viewed in ProteoConnections (<http://www.thibault.irc.ca/proteoconnections>, project 103) and a list of unique phosphosites is provided in **Table A2- I**.

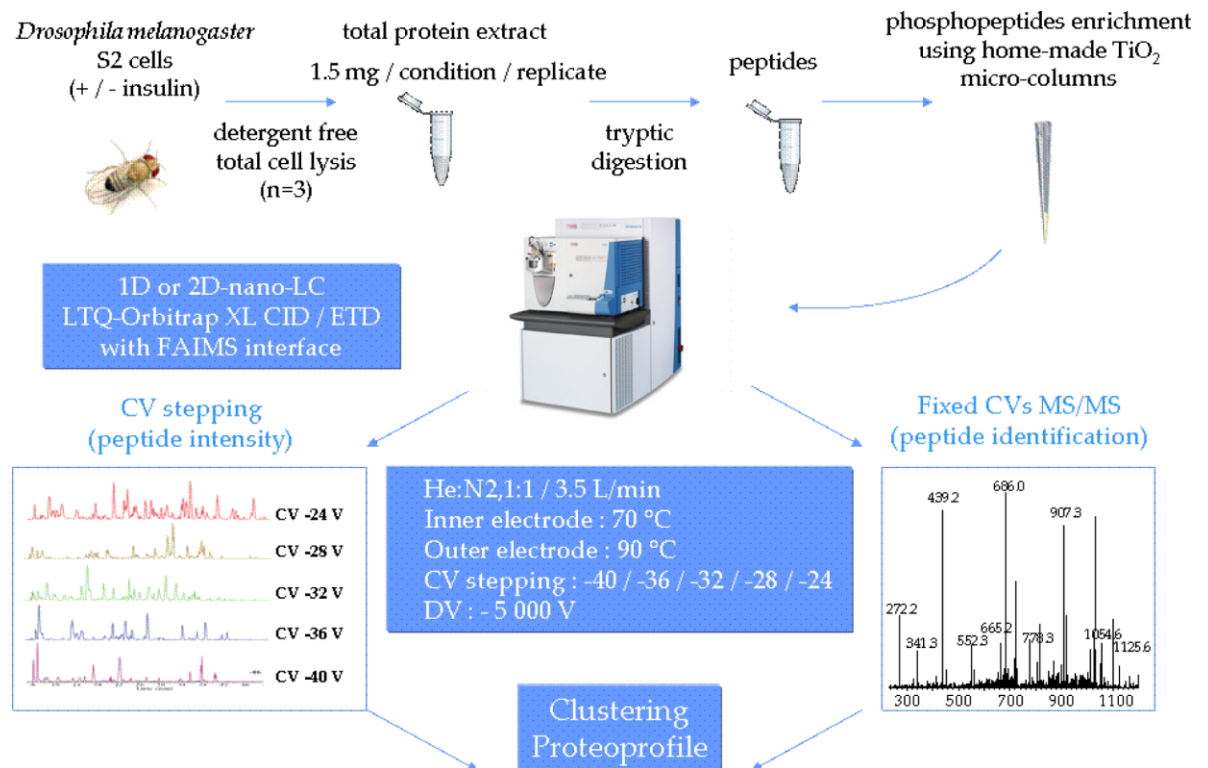


Figure 2.1. Overview of workflow for the comparison of the phosphoproteome of S2 cells with 2D-LC-MS/MS and 2D-LC-FAIMS-MS/MS.

The distribution of identified phosphopeptides and non-phosphopeptides is presented in **Figure 2.2. A** for both approaches. As indicated, a total of 3 476 unique phosphopeptides were identified out of 4 738 peptides using FAIMS compared to 2 168 phosphopeptides from a total of 2 920 peptides without FAIMS. The proportion of phosphorylated and non-phosphorylated peptides across salt fractions remained relatively constant, suggesting that FAIMS or SCX fractionation did not separate these two peptide populations. This was also evidenced in separate infusion experiments on tryptic digest from α 1, α 2 and β -casein where phosphopeptides were transmitted across a relatively large CV domain extending from -45 to -10 V (data not shown). It is noteworthy that successive 2D-LC-MS/MS experiments without FAIMS led to marginal improvements in the number of new peptide identification.

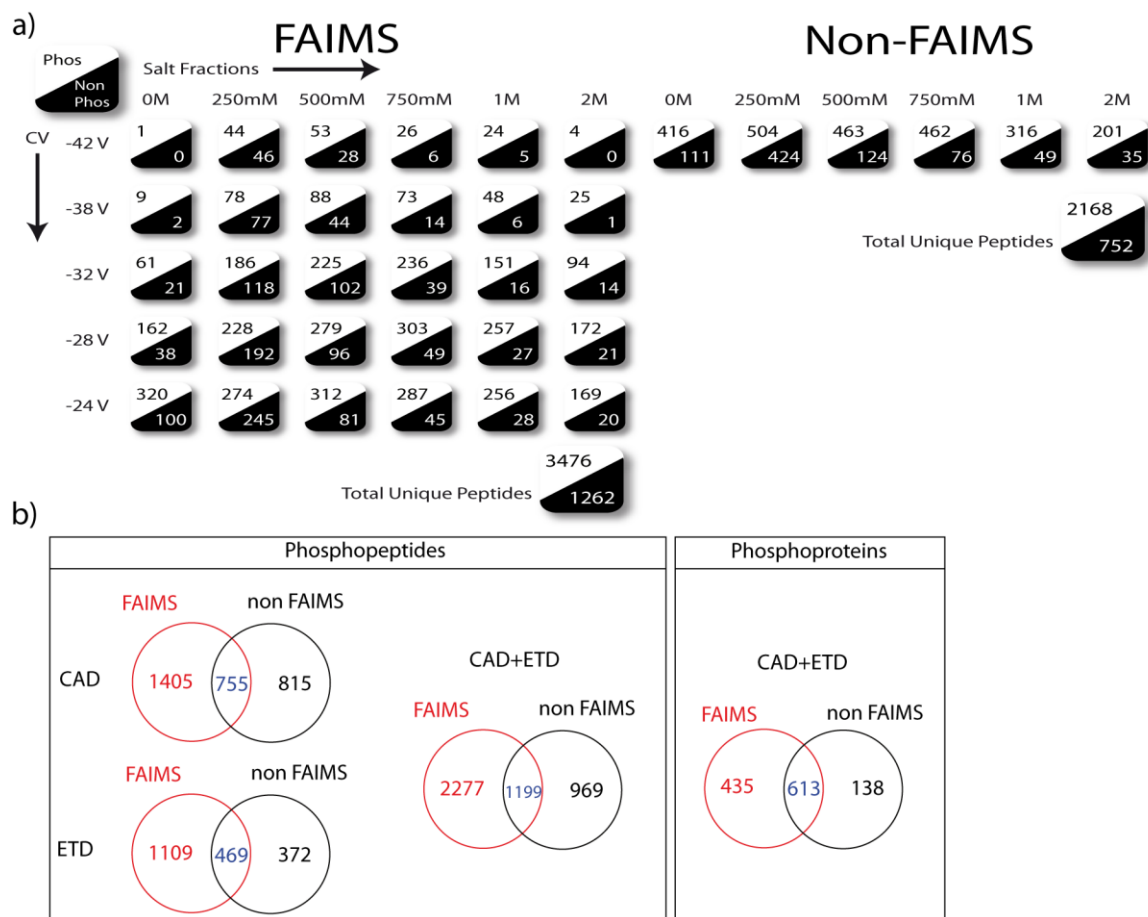


Figure 2.2. Peptides distribution in FAIMS and non-FAIMS experiments.

A. Phospho and non-phosphopeptide distribution across salt fractions and compensation voltages for analyses performed with and without FAIMS. B. Venn diagrams showing the different distribution of phosphopeptide identifications between FAIMS and non FAIMS experiments using ETD or CAD activation.

For example, we observed a gain of 22 and 17% in the acquisition of new peptide ions in the second and third 2D-LC-MS/MS runs, indicating that the proportional increase in newly identified peptides decreased with the number of replicate runs. This result is consistent with previous reports where diminishing returns were observed in the cumulative number of unique peptides beyond three successive injections [326, 327].

The overall distribution of phosphopeptides identified by 2D-LC-MS/MS with and without FAIMS is shown in **Figure 2.2. B** for separate activation modes. In total, we identified 4 445 unique phosphopeptides, of which 67% (2 975 phosphopeptides) were assigned from their corresponding CAD MS/MS spectra. Approximately one third of these spectra (1 199 phosphopeptides) were identified in both FAIMS and non-FAIMS experiments, whereas 51% (2 277 phosphopeptides) were only detected using FAIMS. A similar enhancement in phosphopeptide identification was also observed for ETD MS/MS spectra acquired using FAIMS. In this case, the overlapping phosphopeptides population represented only 24% (469 phosphopeptides) whereas 57% of all ETD spectra (1 109 phosphopeptides) were identified using FAIMS only. While these results highlight that FAIMS facilitates the identification of a distinct subset of the *D. melanogaster* phosphoproteome, a complete overlap with non-FAIMS MS/MS spectra was not possible under the present conditions due to the selection of 5 distinct CV values to transmit ion over the desired CV range. Indeed, phosphopeptides uniquely identified without FAIMS might have optimal ion transmission at CV values intermediate to those selected.

These results indicated that the coupling of FAIMS on the LTQ-Orbitrap XL provided a 49% enhancement in new phosphopeptide identification compared to conventional 2D-LC-MS/MS experiments. To determine if this gain in identification was associated with MS/MS sampling or ion abundance, we profiled the number of assigned phosphopeptides according to precursor intensity for both FAIMS and non-FAIMS experiments (**Figure 2.3**). We correlated the intensity of 3 726 precursor ions from the 4445 identified phosphopeptides shown in **Figure 2.2. B**. As indicated in **Figure 2.3. A**, a comparable distribution of identification was observed between FAIMS and non-FAIMS down to an intensity of 300 000 counts. Larger gains in identification were observed for low-abundance peptide ions where FAIMS enabled the assignment of 1 188 phosphopeptide ions below intensity levels of 300 000 counts compared to 394 without FAIMS.

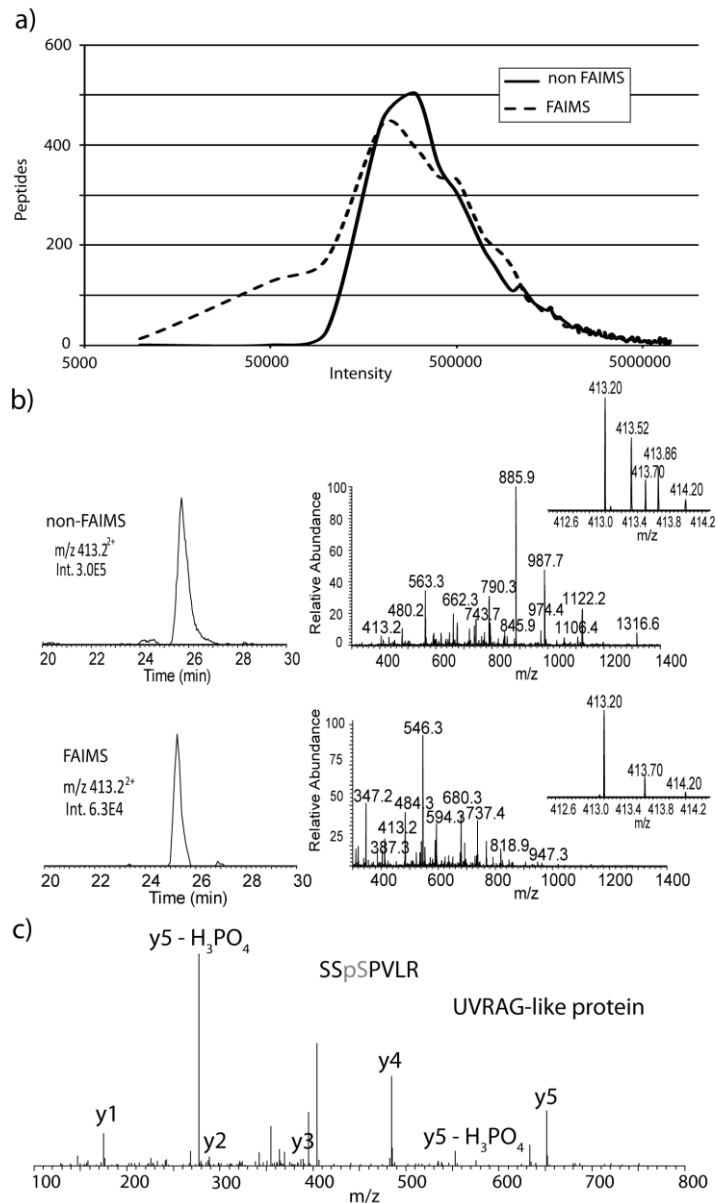


Figure 2.3. Comparison of identified phosphopeptides using FAIMS and non-FAIMS.

A. Intensity distribution of phosphopeptides clusters in FAIMS and non FAIMS experiments showing enhanced identification for low abundance peptide ions using FAIMS. B. Extracted ion chromatograms for the precursor ion at m/z : 413.2²⁺ and extracted mass spectrum of peak eluting at 25.5 min (spectrum with and without FAIMS). C. CAD MS/MS spectrum for precursor at m/z : 413.2²⁺ corresponding to peptide SSpSPVLR from UVRAG-like protein acquired using FAIMS. This phosphopeptide was non identified without FAIMS.

However, the relative increase in phosphopeptide assignment using FAIMS is not only attributed to low abundance peptide ions. Indeed, the successful sequencing of relatively abundant precursors can be compromised by the presence of co-eluting ions with neighboring m/z values. An example of this is presented in **Figure 2.3. B** for the extracted ion chromatogram (XIC) of m/z : 413.2²⁺ obtained using 2D-LC-MS/MS with and without FAIMS. The extracted mass spectra taken at 25.0 min are presented for both analyses and display precursor ions with signal to noise ratio of 20:1 in both cases. While the intensity of ion m/z : 413.2²⁺ is 5-fold higher for the non-FAIMS, the inherent complexity of the mass spectrum observed for this precursor ion prevented its successful identification by MS/MS. In contrast, FAIMS enabled the unambiguous identification of the phosphorylated Ser residue from the UV radiation resistance-associated (UVRAG) like protein using CAD activation (**Figure 2.3. C**). It is noteworthy that we generally observed comparable intensities for peptides identified in both FAIMS and non-FAIMS experiments (**Figure A2. 2**). The observation of peptides with lower intensity in the FAIMS experiment is explained by the non optimal transmission of ions at preset CV values selected for the on-line 2D-LC-FAIMS-MS analysis. Importantly, FAIMS enabled the transmission of low abundance peptide ions with improved S/N ratios and reduced background of confounded noise or co-eluting species to facilitate their identification by MS/MS (**Figure 2.3. A**).

The inherent increase in peak capacity available using FAIMS previously yielded additional advantages in proteomics analyses by reducing the occurrence of MS/MS spectra from mixed precursor ions[233]. In phosphoproteomics, this situation can also arise from co-eluting phosphopeptide isomers that comprise several phosphorylatable Ser/Thr/Tyr residues on a single peptide backbone[328]. Previous reports have described the capability of FAIMS to transmit synthetic phosphopeptide isomers at different CV values[321, 322]. Interestingly, our large-scale *D. melanogaster* phosphoproteomics analyses also revealed the presence of several co-eluting phosphoisomers as evidenced from the variable location of modified residues on the same peptide sequence. Although this observation can be partly explained by the variable confidence in site location, we could not rule out the occurrence of phosphoisomers. In the present experiment, we identified 78 potential phosphoisomers

using FAIMS only. A large proportion of these phosphoisomers (73%) were transmitted at neighbouring CV values. In these cases, the difference in CV is not sufficiently large to prevent the acquisition of mixed MS/MS spectra although different sites were identified by Mascot. However, approximately a quarter of all phosphoisomers were transmitted at CV values at least 10 V apart and enabled the acquisition of clearly distinct MS/MS spectra. The largest differences in CV values were observed for small phosphoisomers where phosphorylated amino acids could potentially form open or more compact structures based on salt bridge with neighbouring basic residues (**Figure A2. 3**). This is illustrated in **Figure 2.4. A** for the monophosphorylated peptide DKSEGDSPNIETPKK from the small-minded protein (CG8571). The ETD MS/MS spectrum acquired using conventional 2D-LC-MS/MS showed a series of low abundance c- and z-type fragment ions, but resulted in an ambiguous assignment with a Mascot score of 3 (**Figure 2.4. A**). The singly-charged fragment ions $z_5 + 1$ and $z_6 + 1$ at m/z : 666.3 and 779.4 suggested the phosphorylation of the Thr residue while c_8 and c_{10} fragment ions at m/z : 833.4 and 1033.5 were consistent with a phosphorylated Ser residue. In contrast, FAIMS enabled the acquisition of these phosphoisomers at two different CV values with distinct peaks separated by only 30 s (**Figure 2.4. B**), while these two phosphopeptides were confounded as a single peak in the non-FAIMS experiment (**Figure 2.4. A**).

The monophosphorylated peptide with a modified Ser7 residue was transmitted at CV -42 V (**Figure 2.4. B**) whereas the Thr12 phosphorylated peptide was detected at CV -28 V (**Figure 2.4. B**). In this case, the complete separation of phosphoisomers using FAIMS enables the unambiguous identification of the corresponding phosphopeptides with distinct fragment ions. For example, the phosphopeptide DKSEGDpSPNIEPTKK was clearly identified with c_{10} and c_{11} fragment ions at m/z : 1140.4 and 1269.5, respectively. Conversely, the phosphopeptide DKSEGDSPNIEPpTKK showed distinct $z_5 + 1$ and $z_6 + 1$ fragment ions at m/z : 666.3 and 779.4 with no cross-talk from the other phosphoisomer. In this particular example, the collisional cross section of these two phosphoisomers is significantly different to favour their transmission at distinct CV values.

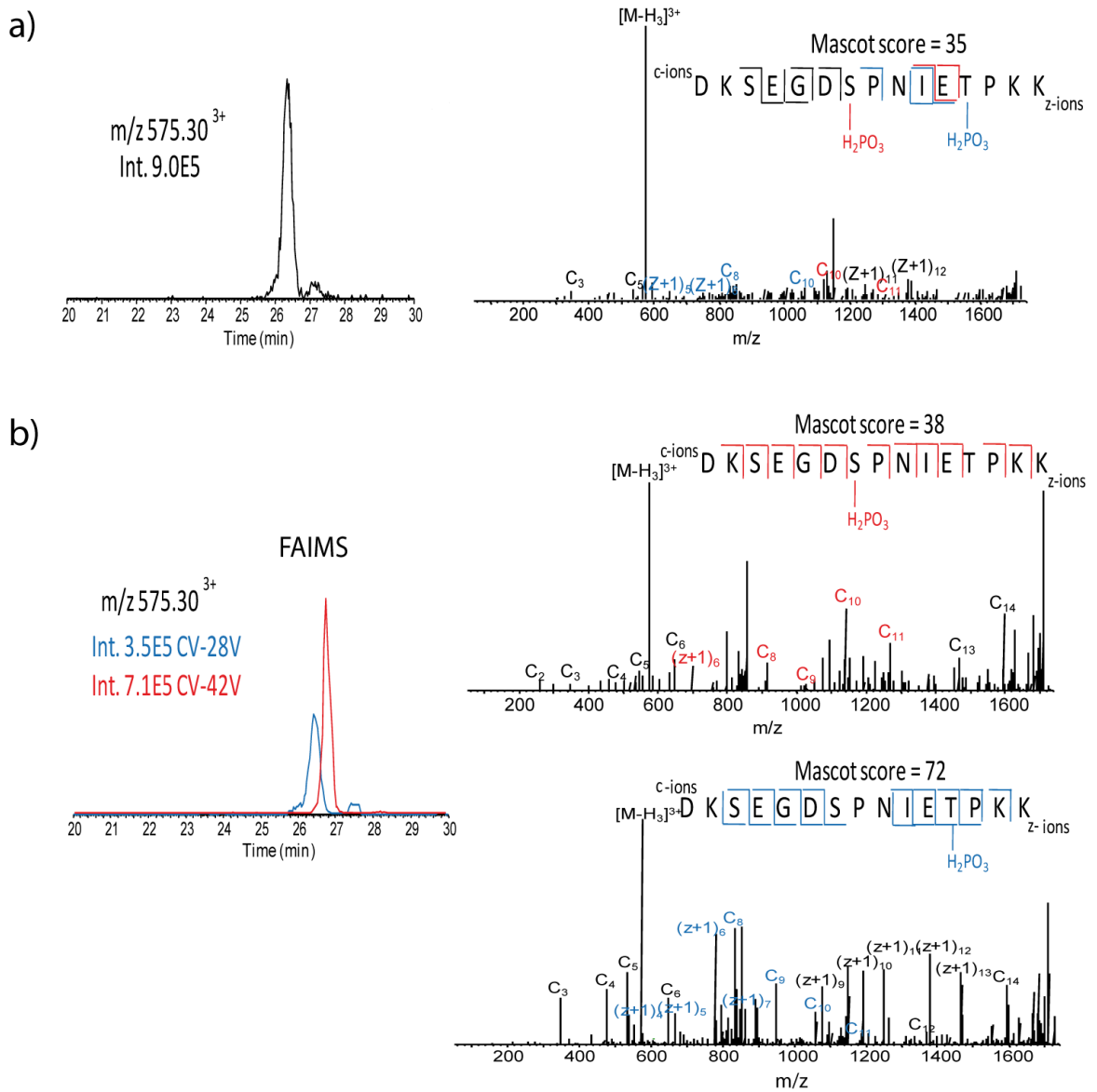


Figure 2.4. Example of phosphopeptide isomers differentiated using FAIMS.

Extracted ion chromatogram and ETD MS/MS spectra for the precursor at m/z : 575.3³⁺ corresponding to peptide DKSEGDSPIETPKK from small-minded protein (CG8571) for 2D-LC-MS/MS analyses performed (A) without FAIMS and (B) with FAIMS at CV: -28 V (blue) and CV: -42 V (red).

2.5.2. Differential phosphoproteome analysis of insulin-stimulated *D. melanogaster* cells

We exploited the advantages of FAIMS in combination with 2D-LC-MS/MS to profile the changes in protein phosphorylation of *D. melanogaster* cells following insulin stimulation. The insulin signaling pathway is highly conserved from drosophila to mammals, and mediates important functions such as carbohydrate and lipid metabolism, cell growth and proliferation [54]. *Drosophila* is a well established animal model system to study insulin and insulin-like growth factor/TOR (target of rapamycin) signaling pathways. In *drosophila*, insulin is known to exert its effects through the activation of several kinases including ERK, PI3K and Akt[329]. To determine the optimal time and concentration of insulin stimulation in *drosophila* S2 cells, we monitored the activation of ERK and Akt by immunoblots (**Figure 2.5. A**). These experiments indicated that ERK is rapidly activated upon insulin stimulation leading to its phosphorylated form (p-ERK) that is progressively dephosphorylated and deactivated 15 min post-stimulation. Similarly, Akt is rapidly activated and remains phosphorylated for at least 60 min. Small changes in the activation of ERK and Akt were observed over the concentration of insulin examined, and a concentration of 0.86 μM with a stimulation time of 8 min were selected for subsequent phosphoproteomics experiments. Independent confirmation of insulin stimulation in *drosophila* S2 cells under these conditions was obtained by monitoring the changes in MAPK phosphorylation using LC-FAIMS-MS. This was achieved by first determining the CV value for maximal transmission of the synthetic peptide IADPEHDHTGFLpTEpYVATR corresponding to the doubly phosphorylated tryptic peptide of MAPK with the activation loop (T₂₄₈/Y₂₅₀). Optimal transmission of this phosphopeptide was obtained at a CV value of -26.5 V. LC-FAIMS-MS analyses were then repeated with tryptic digests of control and insulin-stimulated *drosophila* S2 cells and clearly showed an 8-fold increased in the activated MAPK (**Figure 2.5. B**).

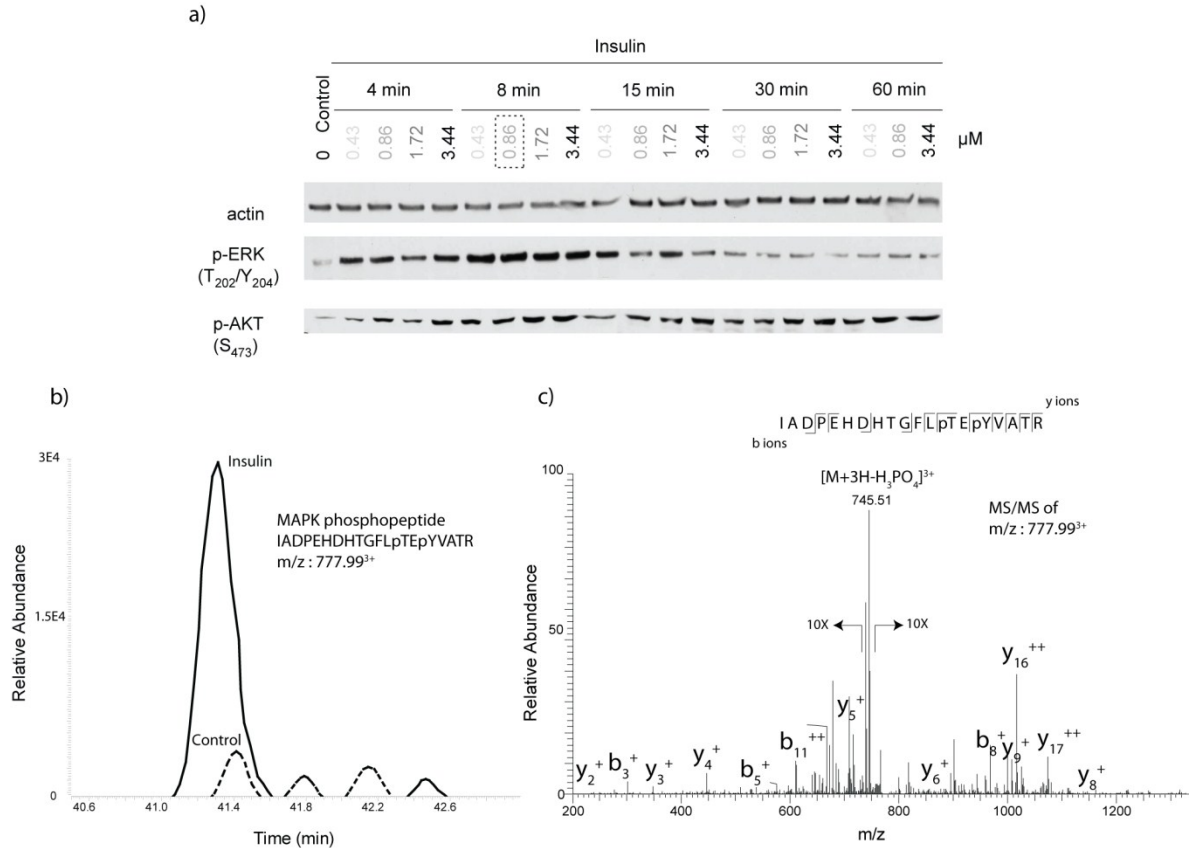


Figure 2.5. Effect of insulin on the phosphorylation of ERK and Akt.

A. Immunoblots of a total cell lysate of S2 cells showing the kinetic of phosphorylation of p-MAPK (T₂₀₂/Y₂₀₄) and p-Akt (S₄₇₃) at different insulin concentrations and incubation times. Insulin rapidly activates MAPK with a maximum phosphorylation after 8 min of cell stimulation followed by a progressive decrease. Akt phosphorylation reached a plateau at approximately 8 min and remained constant up to 60 min post-stimulation. B. Extracted ion chromatograms of MAPK phosphopeptide IADPEHDHTGFLpTEpYVATR (m/z : 777.9943³⁺) showing an increased abundance upon insulin treatment compared to unstimulated cells (8 min stimulation with 0.86 μM insulin). The maximum transmission of m/z : 777.9943³⁺ was obtained at a CV of -26.5 V. C. CAD MS/MS spectrum of IADPEHDHTGFLpTEpYVATR (m/z : 777.9943³⁺).

It is noteworthy that conventional LC-MS analyses of these tryptic digests failed to detect the corresponding phosphopeptide, substantiating again the improvement of sensitivity afforded by FAIMS. Confirmation of phosphopeptide identity was obtained from the LC-FAIMS-MS/MS experiment obtained using CAD activation (**Figure 2.5. C**).

Tryptic digests of control and insulin-activated drosophila S2 cells were enriched in phosphopeptides using TiO₂ columns prior to their analyses using 2D-LC-FAIMS-MS/MS as described previously. Biological triplicates were obtained for each condition. These analyses enabled the identification of 3 953 distinct peptides including 2 660 phosphopeptides that were assigned to 805 unique phosphoproteins. We correlated the abundance of 2 285 phosphopeptides (86% of all identified phosphopeptides) across all sample conditions and replicates. A list of unique phosphosites is presented in **Table A2- II** and these results can also be viewed in ProteoConnections (<http://www.thibault.irc.ca/proteoconnections>, project 145). To identify phosphopeptides showing statistically significant changes in abundance upon insulin stimulation, we performed a two-tail homoscedastic t-test on sample replicates. A volcano plot distribution of p-values vs. fold change is presented in **Figure 2.6** for all quantified phosphopeptides with Mascot score greater than 25. Differentially abundant phosphopeptides showing p-values < 0.1 and fold change greater than 2.7 (7 standard deviations on sample replicates, **Figure A2. 4**), were selected and confirmed by manual validation. We identified a subset of 32 differentially regulated candidates (18 down-regulated and 14 up-regulated). A list of these phosphopeptides is presented in **Table 2-I**.

Two examples of differentially regulated phosphoproteins are shown in **Figure 2.7** for phosphoinositide-3 kinase Dp110 (PI3K92E) and stress-activated map kinase-interacting protein 1 (Sin1). Our phosphoproteomics analysis indicated that the abundance of TDTTSFS**p**SDYELTK from PI3K92E was undetected in unstimulated cells and increase by at least 250-fold upon insulin treatment (**Figure 2.7. A**).

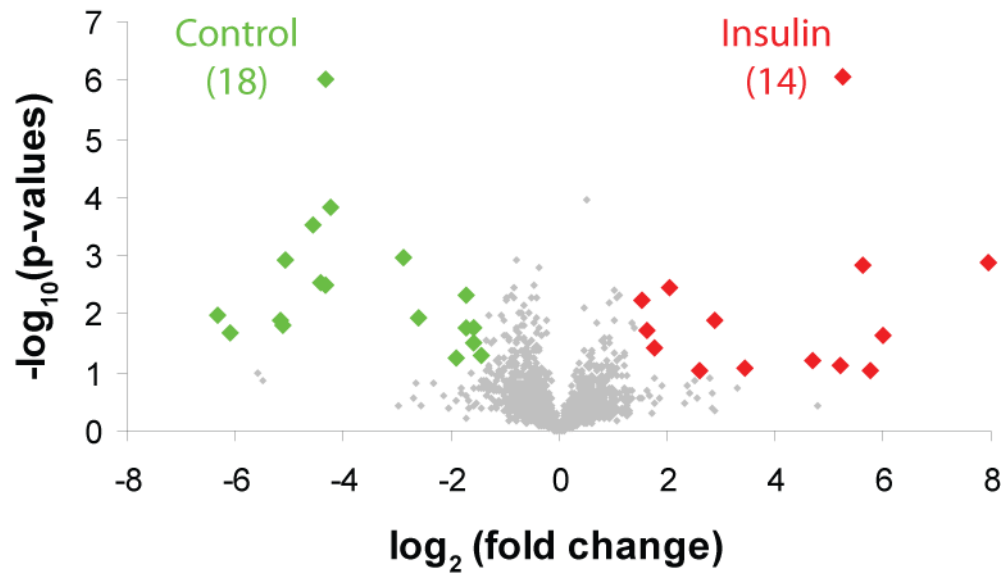


Figure 2.6. Volcano plot distribution of phosphopeptides from control and insulin treated S2 cells.

Only phosphopeptides with a Mascot score greater than 25, and detected in at least 2 replicates were considered. Diamonds correspond to phosphopeptides showing a decrease (green) or an increase (red) in phosphorylation upon insulin treatment (fold-change > 2.7, p-value < 0.1).

Confirmation of the phosphorylation site at Ser136 was obtained from the CAD MS/MS spectrum of precursor ion m/z : 837.9 that showed characteristic y-type fragment ions (**Figure 2.7. B**). PI3K92E is involved in phosphatidylinositol phosphate biosynthesis[330, 331], and is a major target of insulin action[332, 333]. Another example of a phosphoprotein modulated by insulin is Sin1. We observed a 3.5-fold increase in S₁₈₄ phosphorylation upon insulin stimulation (**Figure 2.7. C**) and confirmed the modified residue based on the CAD MS/MS spectrum of its corresponding tryptic peptide (**Figure 2.7. D**). In mammalian cells, Sin 1 isoforms assemble into distinct multi-protein mTORC2 complexes that phosphorylate Akt/PKB. Previous reports have indicated that insulin activates some of these mTORC2 complexes[334]. Whether or not the phosphorylation of Sin1 is required for the assembly of mTORC2 complexes is presently unknown.

Table 2-I. List of differentially phosphorylated peptides after insulin treatment in drosophila S2 cells.

Fold change Control / Insulin	p-value	CG	Associated Gene Name	Peptide Sequence	Phosphosite position	CK2 consensus	Phosphosite conserved in mammals	Function
0.37	0.051	CG5330	Nap1	EPESSpTDNEADAEHFR	T120	yes	no	translation/transcription
0.30	0.005	CG10840	eIF5B	VKESWDApTDpSEAEPEPEEETSQATNNGK	T462 / S464	yes	no/yes	translation/transcription
0.27	0.059	CG2691	CG2691	KKQNLEAQEDpSpSDDELVSGLQEK	S1094 / S1095	yes	no/no	translation/transcription
0.047	0.003	CG4758	Trp1	DEKKDpTDPEGDNGQGDDGASGNEK	T152	yes	no	vesicular trafficking
0.054	0.000	CG6746	CG6746	KILGGSSGpSPQKK	S239	no	no	lipid metabolism
0.049	0.000	CG4225	Hmt-1	NLDAEQSGGpSDNGDASAESGSEK	S824	yes	no	vesicular trafficking
0.028	0.013	CG16742	CG16742	LSIKPTDLFNEFpSDDDTFLSASMSK	S765	no	no	cytoskeleton
0.051	0.033	CG31258	Cenp-C	VNTPTPIDQNSDDVPSRNPpSPSR	S1003	yes	no	translation/transcription
0.051	0.003	CG5899	CG5899	IQVMADpSDpSDGNDSQTPK	S59 / S61	yes	yes/no	translation/transcription
0.16	0.012	CG7962	CdsA	NSAADSpSDHVDpSEEEKIPEEK	S35 / S40	yes	no/no	lipid metabolism
0.30	0.018	CG6395	Csp	KLpSTSGDSLpYEILGLPK	S12	no	yes	vesicular trafficking
0.029	0.016	CG4800	Tctp	LVDDVlpYEVYVK	Y28	no	no	lipid metabolism
0.029	0.001	CG13610	Orct2	VEISDEALSELLDEGENpSEEKAK	S310	yes	no	lipid metabolism
0.33	0.017	CG7989	l(2)k07824	AAWpSDpSDDEDLQVGDVKK	S102 / S104	yes	no/no	translation/transcription
0.042	0.001	CG42236	RanBPM	HHSTPAASNpSPQHQGVDPLR	S395	no	no	translation/transcription
0.14	0.001	CG3456	Met1	STpSNDDIAR	S315	yes	no	lipid metabolism
0.013	0.011	CG8709	CG8709	SDSLVNAENTSALVDNLEELTMpSNKSDEPKER	S748	no	yes	lipid metabolism
0.015	0.021	CG18076	Shot	INDIGAQYDALIRPEpSPAR	S2291	no	yes	cytoskeleton
36.4	0.080	CG15784	CG15784	VSQELIDCDIDFVEDTpTSKVEEEAR	T35	no	no	cytoskeleton
50.2	0.001	CG14561	CG14561	AFSPHGRpTPpSPIDDLSDIETPKQPQR	T69 / S71	no	no/no	-
63.1	0.024	CG4636	SCAR	TPNSMVNEEAGYGNMGVYDTRPPRPNpSIELNR	S262	no	yes	cytoskeleton
3.1	0.019	CG12819	sle	KLpSPQPPVVDKK	S1284	no	no	translation/transcription
54.1	0.094	CG7467	osa	DGAGNGpSDpSEDADEGIDLGQVR	S1930 / S1932	yes	no/no	translation/transcription
4.1	0.004	CG2049	Pkn	ATSpTTPILDQEVIPQLGK	T737	no	no	lipid metabolism
25.8	0.061	CG5641	CG5641	EGDLEEDIDMIENENEEGpSDDGAE	S391	yes	yes	translation/transcription
38.5	0.001	CG7765	Khc	SMFDpSQMDELRL	S724	yes	no	cytoskeleton
10.9	0.087	CG1119	Gnfl	VKEEHTpSPKETK	S376	yes	yes	translation/transcription
6.0	0.091	CG8190	eIF2B-gamma	TVQDTTpSEVGVVTKNEDHVLHYVGHITILDQK	S253	no	no	translation/transcription
3.5	0.037	CG10105	Sin1	ANDDDLpSDEGVQSQLTEQLAK	S184	yes	no	lipid metabolism
7.4	0.013	CG8290	nd	TSSLSDpSMESLESNVSKR	S225	no	yes	-
2.9	0.006	CG11471	Aats-ile	SpSEVLCpVDEAGR	S355	no	yes	translation/transcription
252	0.001	CG4141	Pi3K92E	TDTTSFSpSDYELTK	S136	yes	no	lipid metabolism

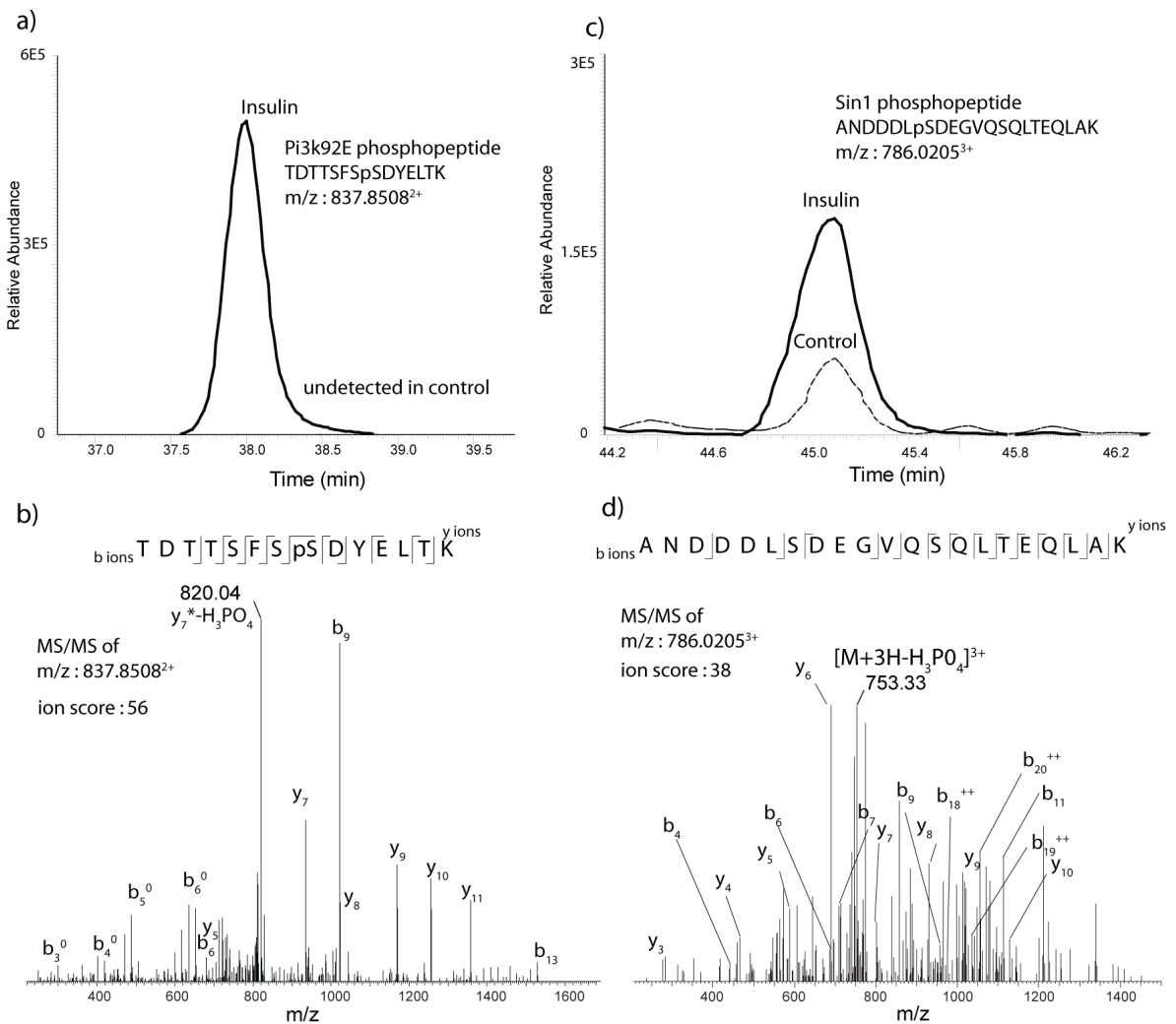


Figure 2.7. PI3K92E (S₁₃₆) and Sin1 (S₁₈₄) are phosphorylated following insulin stimulation.

A. Extracted ion chromatograms (XIC) of m/z : 837.8508²⁺ corresponding to the phosphopeptide TDTTSFSpSDYELTK assigned to PI3K92E (0 mM salt fraction, CV -28 V). B. Tandem mass spectrum of TDTTSFSpSDYELTK using CAD. C. XIC of m/z : 786.0205³⁺ corresponding to the phosphopeptide ANDDDLpSDEGVQSQLTEQLAK assigned to Sin1 (0 mM salt fraction, CV -28 V). D. Tandem mass spectrum of ANDDDLpSDEGVQSQLTEQLAK acquired using CAD.

Phosphopeptides showing differential abundance are regrouped in **Table 2-I** into four categories: translation/transcription, vesicular trafficking, cytoskeleton assembly and lipid metabolism. For instance, Cystein string protein (Csp), a chaperone protein involved in neurotransmitter exocytosis, was previously reported to be phosphorylated on S₁₀ by PKA (S₁₂ for drosophila). Phosphorylation of this residue is required for the binding to 14-3-3, and this interaction might be important to maintain normal synaptic function[335]. We also identified a lipin homologue (GH19076p) and CG6746, two proteins involved in fatty acid synthesis and in the lipid triacylglycerol (TAG) pathway[336]. In both cases, phosphorylation is mediated by mTOR following insulin stimulation[337]. Interestingly, lipin-1 subcellular localization is regulated by its insulin-mediated phosphorylation and interactions with 14-3-3 proteins[338]. The protein translationally-controlled tumor protein (Tctp) is involved in calcium binding and microtubule stabilization. Tctp plays a role in the activation of Ras/Raf/ERK and in the activation of PI3K/Akt pathway[339].

We identified kinesin heavy chain (KHC) (S₇₂₄), CG4636 (Scar) (S₂₆₂) and CG15784 (T₃₅) that all showed increased in phosphorylation upon insulin stimulation. KHC is involved in microtubule binding activity[340], while Scar is a homologue of the WASP/WAVE family in mammals, and is involved in actin polymerization and cytoskeleton reorganization[341]. Finally, different proteins regrouped under the category translation/transcription showed a decrease in phosphorylation upon insulin treatment. These included the nucleosome assembly protein 1 (Nap1) (T₁₂₀), CG31258 (Cenp-C) (S₁₀₀₃), CG5899 (S₅₉/S₆₁), eukaryotic initiation factor 5B (Eif5b) (T₄₆₂/S₄₆₄), U3 small nucleolar RNA-associated protein 18 homolog (l(2)k07824) (S₁₀₂/S₁₀₄), RRP12-like protein (CG2691) (S₁₀₉₄/S₁₀₉₅) and Ran-binding proteins 9/10 homolog (RanBPM) (S₃₉₅). In contrast, Eif2b which mediates a key regulatory step in peptide-chain initiation during mRNA translation, showed an increase in phosphorylation at S₂₅₃. Eif2b was previously reported to be activated by insulin through a signalling pathway involving PI3K and inhibition of GSK-3[342].

We also performed phosphorylation motif analyses to determine consensus sequences that could be over-represented in our datasets. Results from these analyses are

shown in **Figure 2.8** for the most significant motifs identified by Motif-X. Consensus sequences representing more than 5% of identified phosphopeptides are regrouped in **Figure 2.8. A**. The six most significant groups are represented by the motifs: SXE, SP (ERK1/ERK2), EXXXXS, DXXXXS, DS and SDXD (where X corresponds to any amino acid). Interestingly, when the same analyses were performed using only phosphopeptides regulated by insulin (**Table 2-I**), we observed that almost half of these sequences were represented by the motif SXXXD corresponding to casein kinase 2, CK2 (**Figure 2.8. B**). This observation suggests that CK2 is activated upon insulin stimulation. Interestingly, previous reports indicated that insulin induced a rapid activation of MAPK and CK2 in 3T3 mouse adipocytes, and a selective translocation of MEK from the cytoplasm to the nucleus[127, 343]. CK2 is a highly conserved and ubiquitous kinase amongst eukaryotes, and is involved in several intracellular pathways regulating cell cycle, proliferation, apoptosis and transformation. CK2 subunits are known to interact with Akt, an association that enhances the activity of the latter kinase. While these results suggest that insulin activate CK2, the nature of substrates phosphorylated by this kinase is presently unknown and warrant further investigations.

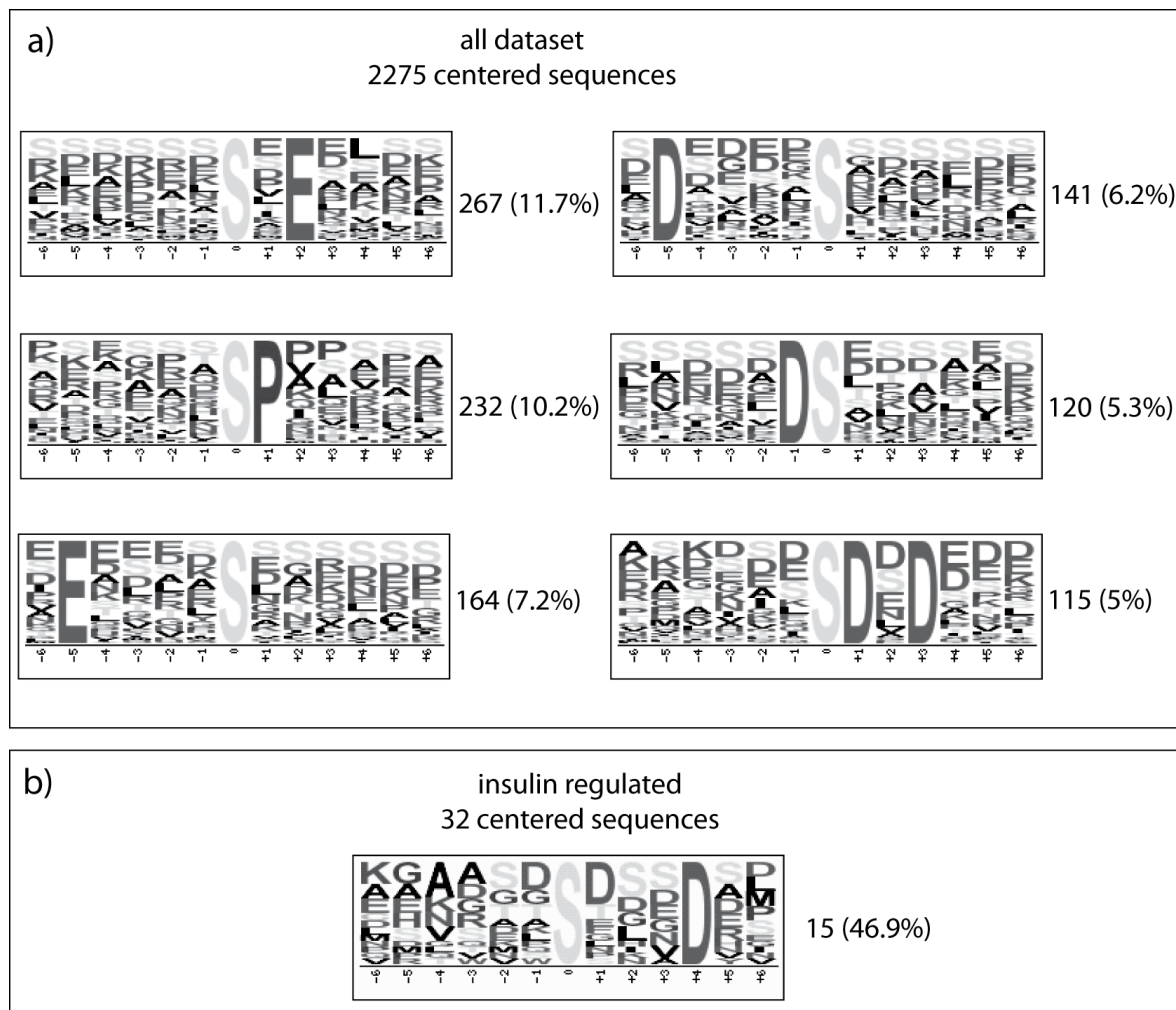


Figure 2.8. Phosphorylation motifs for phosphopeptides identified in insulin treated S2 cells.

A. Distribution of the most abundant phosphorylation motifs identified in large-scale phosphoproteome analyses of *D. melanogaster* S2 cells stimulated by insulin. B. Phosphorylation motif represented by phosphopeptides regulated by insulin treatment. Almost half of insulin-regulated phosphopeptides have a SXXXD motif (potential CK2 consensus motif).

2.6. Conclusion

The identification of phosphopeptides depends not only on the abundance of precursor ions, but also on the occurrence of co-eluting ions of similar m/z values that could complicate the interpretation of the corresponding MS/MS spectra. In this context, FAIMS plays an important role in phosphoproteome analyses by reducing the contribution of contaminant ions (e.g. column bleed, solvent clusters and other isobaric ions) and enhancing the peak capacity of LC-MS/MS. Although these advantages are relevant for any large-scale proteome analyses, their significance might be more important in phosphoproteomics where ionization suppression effects, low abundance precursor ions and isomeric ions can compromise the successful identification of phosphopeptides. The coupling of FAIMS to 2D-LC-MS/MS provided marked advantages for MS/MS sequencing by enriching the proportion of multiply charged phosphopeptides while minimizing the contribution of undesired background ions or co-eluting phosphopeptides. This was clearly evidenced by a 49% increase in the number of identified phosphopeptides from the S2 phosphoproteome using FAIMS compare to conventional 2D-LC-MS/MS experiments. FAIMS also provides the ability to separate phosphopeptide isomers that cannot be resolved by LC alone and could result in ambiguous MS/MS assignments. This advantage is of relevance in phosphoproteomics to correctly assign sites modulated by specific cell stimulus especially when different phosphorylatable residues are present on the same tryptic peptide. The analytical merits of FAIMS in large-scale phosphoproteomics were exploited to identify *D. melanogaster* proteins showing rapid changes in phosphorylation upon insulin stimulation. We profiled changes in the abundance of 2 660 phosphopeptides, and identified a subset of 32 proteins involved in translation/transcription, vesicular trafficking, lipid metabolism and cytoskeleton that were regulated by insulin. Among these, we confirmed the changes in phosphorylation of MAPK and PI3K92E previously known to be phosphorylated in response to insulin treatment. Importantly, close to 50% of phosphoproteins regulated by insulin displayed a consensus motif for CK2, suggesting that this kinase is rapidly activated upon insulin stimulation.

2.7. Acknowledgements

IRIC receives infrastructure support funds from the Fonds de la Recherche en Santé du Québec (FRSQ). This work was carried out with the financial support of operating grants from the National Science and Engineering Research Council (NSERC) and the Canada Research Chair programs. The authors also thank the following colleagues from Thermo Juergen Srega and Brenda Kesler for insightful suggestions and logistic support, and Michael Belford for technical assistance.

Chapitre 3 : A combined phosphoproteomics and RNAi approach to identify CK2 substrates following insulin stimulation of Drosophila S2 cells

Gaëlle Bridon^{1,2}, Éric Bonneil¹, Olivier Caron-Lizotte¹, Pierre Thibault^{1,2,3}

Manuscrit en préparation pour *Journal of Proteome Research*

¹Institute for Research in Immunology and Cancer, ²Department of Biochemistry, Université de Montréal, P.O. Box 6128, Station. Centre-ville, Montréal, Québec, Canada H3C 3J7, ³Department of Chemistry, Université de Montréal.

3.1. Contribution des auteurs

Gaëlle Bridon et Pierre Thibault ont écrit le manuscrit. Éric Bonneil a révisé et commenté le manuscrit. **Gaëlle Bridon** a effectué la culture cellulaire, la préparation des échantillons, l'analyse par spectrométrie de masse ainsi que le traitement des données. Éric Bonneil a participé à la mise en place de l'instrumentation. Olivier Caron-Lizotte a programmé le logiciel Proteoprofile servant à l'analyse bioinformatique des données. Pierre Thibault a supervisé cette analyse.

3.2. Abstract

This study reports the use of label-free quantitative phosphoproteomics to profile changes in protein phosphorylation of *Drosophila melanogaster* S2 cells stimulated by insulin. High field asymmetric waveform ion mobility spectrometry (FAIMS) combined to LC-MS/MS on a LTQ-Orbitrap Velos was used to enhance phosphoproteome coverage and to determine differentially phosphorylated proteins following insulin stimulation. These analyses enabled the identification of 1 569 sites including 51 sites that were modulated by insulin treatment. Approximately 40% of differentially regulated phosphorylation sites contained a potential CK2 consensus motif, of which 15 were confirmed by large-scale phosphoproteomics and RNAi experiments. Interestingly, we identified six proteins involved in chromatin organization and DNA damage response that were regulated by both insulin and CK2. *In vitro* kinase assays using synthetic peptides confirmed CK2 specificity and identified residue Ser15 of heterochromatin protein 1 as a direct CK2 substrate also modulated by insulin stimulation.

3.3. Introduction

Protein phosphorylation is a reversible post-translational modification (PTM) that regulate proteins function and play a significant role in a wide range of cellular processes such as cell division, growth and differentiation through complex signaling pathways. The concerted action of kinases and phosphatases can impart significant conformational changes on many enzymes and receptors to regulate their activities. Phosphorylation of a specific residue can also modulate the interaction of a given protein with other binding partners or affect its cellular location. The interplay between phosphorylation and other proximal PTMs (e.g. ubiquitylation[344], SUMOylation[345]) can modulate the function of proteins involved in regulatory networks. To profile the dynamic changes in protein phosphorylation and their impact on the regulation of different biological activities, analytical methods must provide an appropriate level of sensitivity and comprehensiveness. Recent advances in both affinity chromatography (IMAC[307-310], TiO₂[19, 142, 311], SIMAC[171] or PolyMAC[180]) and sensitive mass spectrometry (MS) instruments now facilitate the profiling of thousands of phosphopeptides from complex cell extracts in response to different cell stimuli.

Furthermore, enhancement of MS sensitivity for proteomics and phosphoproteomics analyses can be achieved using high field asymmetric waveform ion mobility spectrometry (FAIMS)[1, 232, 233]. The use of FAIMS was recently shown to provide a 50% increase in phosphopeptide identification compared to conventional LC-MS/MS. FAIMS facilitates the separation of gas-phase peptide ions based on their difference in mobility at high vs. low electric fields[346]. This feature was previously exploited in LC-MS/MS experiments to transmit multiply-charged peptide ions against a background of singly-charged contaminants thus enhancing both sensitivity and signal to noise for proteomics analysis[233, 234]. Moreover, FAIMS provides partial separation of isobaric peptides and phosphopeptides according to their CV values to increase confidence in assignment[321, 322].

This instrumental strategy is applied here to study insulin signaling in *Drosophila melanogaster* S2 cells. Insulin signaling plays several roles such as the control of lifespan, reproduction or metabolism regulation as revealed by various genetic analysis in *drosophila*[48, 49]. The insulin-like receptor (InR) is activated upon insulin binding[51], resulting in its autophosphorylation and leading to the recruitment of the insulin receptor substrates Chico (Irs-1) and Lnk[60, 61]. InR belongs to the family of the tyrosine kinase receptors involved in many types of cancers[29] and the phosphorylation cascade involved downstream is well conserved across evolution. The activation of InR results in the activation of Ras and ERK (extracellular signal-regulated kinase)[53, 55] where Ras-GTP binds to and activates the catalytic subunit of PI3K[347] which leads also to the recruitment of PDK1, AKT and TOR via a relay of phosphorylation events[348]. Highlighting the molecular connections between insulin signaling components and cellular enzymes is essential to understand how insulin acts in S2 cells. Phosphoproteomics studies have already been conducted to dissect the insulin signaling pathway in brown[81] or 3T3-L1[80] adipocytes, mouse liver cell lines[84] and identify proteins from the MAPK pathway[349] or in complex with Chico[83].

Using a differential phosphoproteomics approach in a previous publication, we reported 32 differentially phosphorylated proteins in S2 cells after insulin stimulation[1]. We highlighted phosphoproteins involved in translation/transcription, vesicular trafficking, cytoskeleton assembly and lipid metabolism to be modulated by insulin. We further hypothesized that casein kinase 2 (CK2) could be involved in insulin signaling as we showed that almost 50% of the phosphopeptide hits regulated by insulin contain the S*/T*XXD/E motif that correspond to the CK2 consensus site[93] (according to Motif-X software). CK2 is a ubiquitous serine/threonine protein kinase constituted of a heterotetramer formed by 2 catalytic subunits (α) and 2 β regulatory subunits required for substrate recognition, catalytic activity and stability[89-92]. In *Drosophila melanogaster*, one CK2a and three CK2b subunits (DmCK2b, DmCK2b0 and DmCK2btes) have been described as well as their tissues-specific distribution[131-133]. CK2 is a highly conserved kinase in eukaryotes and is known to phosphorylate hundreds of substrates. It is implicated

in cell growth, proliferation[91], cell cycle progression and cell viability[94, 95]. It was shown to be constitutively active[122, 123] and studies revealed an increase in CK2 activity in cell culture following serum[125], insulin or IGF-1 treatments[126, 127], and EGF stimulation[128], but these conclusions are still discussed[129].

The present paper is a continuation of what has been discovered in our previous publication[1]. We hypothesised that CK2 is involved in the insulin pathway and we conducted a global differential phosphoproteomics experiment using LC-FAIMS-MS/MS in combination with a RNAi approach to inhibit CK2 expression. The goal of this study is to identify phosphorylation events occurring in the insulin signaling pathway in drosophila S2 cells that are modulated after CK2 inhibition. We want to highlight insulin- and CK2-dependent targets.

3.4. Materials and methods

3.4.1. Materials

Modified porcine sequencing grade modified trypsin was obtained from Promega (Madison, WI, USA). Acetonitrile (ACN) and HPLC grade water was purchased from Fisher Scientific (Whitby, ON, Canada). Ammonium bicarbonate and formic acid were obtained from EM Science (Mississauga, ON, Canada). Ammonium hydroxide, trifluoroacetic acid (TFA), DTT (DL-dithiothreitol), iodoacetamide, sucrose, MgCl₂, CaCl₂, Hepes, NaCl, glycerol, protease inhibitor cocktail (4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), pepstatinA, E-64, bestatin, leupeptin, and aprotinin), phosphatase inhibitor cocktail (sodium vanadate, sodium molybdate, sodium tartrate, and imidazole), serum-free and protein-free insect medium-1, human insulin solution and propionic anhydride were purchased from Sigma-Aldrich (Oakville, ON, Canada). Bradford protein assay and Criterion XT precast gels (4-12% Bis-Tris) were obtained from Bio-Rad (Mississauga, ON, Canada). Lactic acid was purchased from Fluka (Saint Louis, MO). Tris

was purchased from EMD Omnipur (Lawrence, KS). Titanium dioxide (Titanium Bulk, 5 μm , 500 mg) was obtained from Canadian Life Science (Peterborough, ON). Bond breaker TCEP (Tris[2-carboxyethyl] phosphine) was purchased from Pierce Biotechnology Inc. (Rockford, IL). Phosphate buffered saline (PBS) and ethylene-diamine-tetra-acetic acid (EDTA) were obtained from HyClone (Thermo Scientific, Logan, UT). Oasis HLB cartridges (1cc, 30 mg) were purchased from Waters (Milford, MA). ECL chemiluminescence detection system was purchased from Amersham Pharmacia Biotech (Montréal, QC, Canada). Actin, p-ERK1/2 (T₂₀₂/Y₂₀₄) and p-AKT (S₄₇₃) primary antibodies were purchased from Cell Signaling (Danvers, MA). Goat anti-mouse IgG and goat anti-rabbit IgG secondary antibody were purchased from Millipore (Ottawa, ON, Canada). Jupiter C₁₈ (3 μm , 300 Å pore size) particles from Phenomenex (Torrance, CA) and fused silica tubing from Polymicro Technologies (Phoenix, AZ) were used to pack capillary columns for LC-MS. Recombinant protein kinases CK1 and CK2 and synthetic peptides (CK1SP and CK2SP) were purchased from New England Biolabs (Pickering, ON, Canada). The set of 132 synthetic peptides containing a Lys or Arg residue in C-ter was purchased from JPT Peptides Technologies (Berlin, Germany).

3.4.2. Cell culture and insulin treatment

Culture conditions were described previously[1] and are summarized below for convenience. *Drosophila melanogaster* Schneider S2 cells were cultured in serum-free and protein-free insect medium at 28 °C. Cells were plated in 100-mm Petri dishes at a density of 18 millions cells per dish 24 hours before treatment. For insulin stimulation, cells were treated for 8 min with an insulin concentration of 0.86 μM .

3.4.3. RNAi synthesis and treatment

PCR were conducted using fly genomic DNA template and Taq polymerase enzyme. Sequences for GFP primers are (5'→3')

GAATTAATACGACTCACTATAGGGAGACGTAAACGGCCACAAGTTCAG and (3'→5') GAATTAATACGACTCACTATAGGGAGAACGAACTCCAGCAGGACCATG. Sequences for CK2-alpha primers are (5'→3') GAATTAATACGACTCACTATAGGGAGAGACACTTCCTAGTGCGGCTCGCGTG and (3'→5') GAATTAATACGACTCACTATAGGGAGAGTAATCATACTCTGGTAATCTACC. All primers sequences contained a specific reverse sequence corresponding to the T7 polymerase for the subsequent reverse transcription. Primers sequences have been designed using the SnapDragon program to prevent off-targets effects[350, 351]. Reverse transcription reaction was incubated overnight at 37°C using 100 mM NTP and 50 U of T7 polymerase (Invitrogen). RNA product was isolated using a phenol/chloroform extraction and ethanol/sodium acetate precipitation. Single strands RNA were annealed after boiling and cooled down to form double strand RNA after hybridation. S2 cells were treated with 15 µg/mL of dsRNA (GFP negative control or CK2-alpha) and incubated during 5 days. Inhibition was assessed by Western blot and RT-q-PCR.

3.4.4. Immunoblotting analysis

Total protein extracts were separated by SDS-PAGE (4-12%) prior to electroblotting onto a PVDF membrane. After blocking non-specific binding sites with 5% non-fat milk, membranes were incubated with the primary antibody (p-ERK1/2 [T₂₀₂/Y₂₀₄], p-AKT [S₄₇₃], actin or CK2-alpha) followed by horseradish peroxidase-conjugated secondary antibody. The secondary antibody was detected with the ECL chemiluminescence detection system.

3.4.5. RT-q-PCR

Total RNA was extracted using TRIzol reagent (Invitrogen). For RT-q-PCR, 500 ng of total RNA was reversely transcribed using the High Capacity cDNA Archive Kit with random primers (Applied Biosystems). PCR reactions were performed using the TaqMan fast Universal PCR Master Mix (Applied Biosystems), specific primers and the Universal

TaqMan probe. The ABI PRISM 7900HT Sequence Detection System (Applied Biosystems) was used to detect the amplification level.

3.4.6. Protein extraction

Cells were washed with PBS and total cell protein extraction was performed using a detergent-free cell lysis in extraction buffer (20 mM Hepes, 1.5 mM MgCl₂, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol) using a dounce metal homogenizer. After spinning at 13 000 g, the supernatant was kept and the membrane pellet was removed. During buffer preparation, 1mM DTT, proteases and phosphatases inhibitor cocktails were added. Total protein amount was quantified using Bradford protein assay.

3.4.7. Trypsin digestion

Proteins in extraction buffer were reduced in 50 mM ammonium bicarbonate containing 0.5 mM TCEP and 1% SDS for 20 min at 37°C, and then alkylated in 5 mM iodoacetamide for 20 min at 37°C. Finally, 5 mM DTT was added to the protein solution to react with excess iodoacetamide. After dilution with 50 mM ammonium bicarbonate, protein digestion was performed overnight at 37 °C in 0.1% SDS with sequencing grade modified trypsin (enzyme/protein ratio: 1/50). The digest mixture was acidified with TFA and then evaporated to dryness in a SpeedVac.

3.4.8. TiO₂ phosphopeptides enrichment

For the TiO₂ enrichment procedure, all steps were performed by spinning the micro-column at 8 000 rpm at 4°C in a regular Eppendorf microcentrifuge as described in Bridon, *et al.*[1]. Briefly, tryptic peptides were resuspended in 250 mM lactic acid (3% TFA/70% ACN) before loading on the TiO₂ column. Six mg of titanium dioxide stationary phase

were packed in a 200 μ L tip to enrich 750 μ g of protein digest. Columns were previously equilibrated with 3% TFA/70% ACN before sample loading. Each column was then washed with lactic acid solution followed by 3% TFA/70% ACN to remove non-specific binding of non-phosphopeptides. Phosphopeptides were eluted with 1% NH_4OH pH 10 in water and acidified with 3% TFA before desalting on Oasis HLB cartridges. Samples were dried in a SpeedVac and the equivalent of 1.5 mg of starting material (amount estimated before phosphopeptide enrichment) was resuspended in 65 μ L of 0.2% formic acid/5% ACN for 1D-LC-FAIMS-MS/MS experiments on the LTQ-Orbitrap Velos mass spectrometer. Samples were prepared in biological triplicates for each condition.

3.4.9. *In vitro* kinase assays to confirm CK2 substrates

In vitro kinase assays were performed using recombinant casein kinase 1 (CK1) and casein kinase 2 (CK2) alpha-subunit purchased from New England Biolabs, NEB (Pickering, ON, Canada). CK1 and CK2 substrates (CK1SP, RRKDLHDDEEDEAMSITA and CK2SP, RRRDDDSDDD-amid) were obtained from NEB. All synthetic peptides (9 pmoles each) were incubated together during 30 min at 28°C using a 1/1 molar ratio of kinase/substrate. CK1 reaction buffer composition was 50 mM Tris-HCl, 10 mM MgCl_2 and 5 mM DTT (pH 7.5) and CK2 reaction buffer contained 20 mM Tris-HCl, 50 mM KCl and 10 mM MgCl_2 (pH 7.5). Kinase reactions were performed with 1 mM ATP in a final volume of 60 μ L and shaking at 600 rpm. Phosphorylation kinetics were determined by sampling 5 μ L of peptide reaction mixture (750 fmoles each) quenched with 20 μ L of 0.2% FA and immediately frozen in liquid nitrogen. Samples were analyzed by 1D-LC-MS/MS on a LTQ-Orbitrap XL at different time intervals (0, 5, 10, 15, 20, 25 and 30 min). No phosphopeptide enrichment was performed before MS analysis as the amount of recombinant protein kinase was sufficient to yield detectable phosphorylated substrate following 30 min incubation.

3.4.10. Casein kinase 2 substrate peptide (CK2SP) propionylation

Because of its high hydrophilicity and poor retention on a C₁₈ column, the N-terminus of CK2SP (RRRDDDSDDD-amid) was propionylated by incubating 50 µL of 10⁻³ M stock solution (5x10⁻⁸ mol) with 100 µL of propionic anhydride and 150 µL of water at room temperature with shaking during 30 min. Sample was dried in a SpeedVac.

3.4.11. *In vitro* kinase assay to evaluate CK2 activity

Endogenous CK2 activity was assessed using CK2 substrate peptide (prop-RRRDDDSDDD-amid). S2 cells (control and treated with insulin or RNAi) were lysed in 10 mM K₂HPO₄ pH 7.5, 1 mM EDTA, 10 mM MgCl₂, 5 mM EGTA and 0.5% NP40 supplemented with protease and phosphatase inhibitor cocktails. After sonication, cell extract was centrifuged at 13 000 g for 15 min at 4°C and the supernatant was kept. The kinase reaction was performed according to Kubota *et al.*[352] using 24 µg of S2 cell lysates and 8 nanomoles of CK2SP in a final volume of 50 µL with 5 mM ATP in the reaction buffer (25 mM Tris pH 7.5, 7.5 mM MgCl₂, 0.15 mM EGTA and 0.1 mM DTT supplemented with protease and phosphatase inhibitor cocktails). The phosphorylation kinetics was determined by sampling 10 µL of reaction mixture (1.6 nanomoles) quenched with 1 µL of TFA looking at 4 times points (0, 15, 30 and 60 min). Samples were dried in a SpeedVac and phosphopeptide enrichment was performed using TiO₂ stationary phase. Eluates were analyzed by LC-MS/MS on a LTQ-Orbitrap XL.

3.4.12. Liquid chromatography

Nano-LC separation was performed using an Eksigent nano-LC (Dublin, CA) system equipped with a Finnigan AS autosampler (Thermo Fisher Scientific) and coupled to a LTQ-Orbitrap XL or Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Custom reversed-phase trap (4 mm length, 360 µm i.d.) and analytical columns (10 cm

length, 150 μm i.d.) were packed with Jupiter C₁₈ stationary phase (3 μm particles, 300 Å pore size, Phenomenex, Torrance, CA). Chromatographic separations of peptides were typically achieved using a flow rate of 600 nL/min and a linear gradient of 5-40% ACN (0.2% formic acid) over 70 min.

3.4.13. FAIMS interface and mass spectrometry

The FAIMS interface was coupled to a LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) via a nanoelectrospray source. The FAIMS electrodes were operated at a dispersion voltage of -5 000 V with a 3.5 L/min flow of helium/nitrogen 1/1. The temperature of the inner and outer electrodes was set to 70 and 90°C, respectively. The Ion Max source was modified for nanoelectrospray ionization. For CV stepping FAIMS experiments, the Xcalibur software enables a programmable CV function to be synchronized with the operation of the LTQ-Orbitrap Velos. CV stepping runs consist of cycling through 5 CV values at a rate of 1s/CV step with 0.3 s interscan time while the corresponding data for each CV is acquired by the mass spectrometer. The conventional MS spectra (survey scan) were acquired at high resolution ($M/\Delta M$: 60 000 full width half maximum, AGC: 3×10^5 , injection time: 300 ms) over the acquisition range of m/z 300-2000. HCD MS/MS spectra were collected using data-dependent acquisition (DDA) for multiply charged ions with an intensity higher than 4 000 counts (isolation width: 2 Da, normalized collision energy: 40, activation time: 20 ms, dynamic exclusion: 30 s). In each cycle, 6 MS/MS scans on the most abundant peptides were performed. For non-FAIMS experiments on the LTQ-Orbitrap XL, the survey scans were acquired at a resolution of 60 000 (AGC: 10^6 , injection time: 500 ms). MS/MS spectrum were acquired in CID mode using DDA for multiply charged ions exceeding a threshold of 10 000 counts (isolation width: 2 Da, normalized collision energy: 35, activation q: 0.25, activation time: 30 ms, dynamic exclusion: 30 s).

3.4.14. Database searching and phosphoproteomics analyses

MS data were analyzed using the Xcalibur software (version 2.0.7). Peak lists were generated using the Mascot Distiller software (version 2.3.2.0, Matrix science) where MS processing was achieved using the LCQ_plus_zoom script. Database searches were performed using the search engine Mascot (version 2.2.0, Matrix Science, London, UK) and Uniprot *Drosophila melanogaster* (Dmel) forward/reverse database containing 58 904 entries (version 3.54). For convenience, a specific database was created to process synthetic peptides experiments. For HCD runs, the error window for experimental peptide mass values and MS/MS fragment ions were set to +/-10 ppm and 0.02 Da, respectively. For CID runs, MS/MS fragment mass tolerance was set to 0.5 Da. The number of allowed missed cleavage sites for trypsin was set to 2 and phosphorylation (STY), oxidation (M), carbamidomethyl (C), and deamidation (NQ) were selected as variable modifications. No fixed modification was included in the search. A cut-off score threshold was established for a false-discovery rate of less than 2%[323]. All peptide identifications were transferred to ProteoConnections, a bioinformatics platform facilitating the organization and interpretation of proteomics data[324]. The confidence in the location of phosphorylation sites was determined using the A-Score based on the probability of a specific assignment for fragment ions described by Olsen *et al.*[19]. Phosphorylation sites were also annotated according to consensus motifs (Motif-X[9]), kinase prediction tools (NetworKIN[294] and PPSP[353]) integrated within ProteoConnections. Identified phosphosites were centered with six flanking amino acids on each side of the modified residue. Parameters set for the width, occurrences, significance and background were 13, 20, 0.000001 and Flybase + *Drosophila* Proteome, respectively.

3.4.15. Peptide detection, clustering and validation

The comparison of peptide abundances across different conditions and replicates was achieved using label free quantitative proteomics[233, 325] as described in Bridon, *et al.*[1]. Briefly, raw data files (.raw) from the Xcalibur software were converted into peptide

map files representing all ions according to their corresponding m/z values, retention time, intensity and charge state with a defined intensity threshold (determined from background values). Peak top values were considered as peptide intensity and only the CV allowing the maximal ion transmission was considered to calculate peptide intensity for FAIMS experiments. Clustering parameters were set to ± 15 ppm and ± 1 min for peptide mass and retention time, respectively. Reproducible changes in abundance across conditions were determined using a two-tailed homoscedastic t-test on sample replicates to identify peptide clusters with p-values < 0.1 and fold changes greater than 7 standard deviations. Peptide clusters fulfilling these selection criteria were inspected manually to validate identification and changes in abundance. Global intensities were normalized based on the average intensity of all peptide signals in 3 replicates.

3.5. Results and discussion

The primary objective of the present study was to profile changes in protein phosphorylation in response to insulin stimulation, and to identify differentially regulated proteins using 1D-LC-FAIMS-MS/MS (**Figure 3.1**). Preliminary experiments performed using label-free phosphoproteomics on *Drosophila melanogaster* S2 cells suggested that a large proportion of differentially regulated phosphosites harboured a casein kinase 2 (CK2) consensus motif[1]. Further to these experiments, we isolated drosophila protein extracts from control and insulin treated cells with RNAi knock-down for CK2 and GFP. The corresponding cell extracts were lysed, digested with trypsin and phosphopeptides were enriched on TiO_2 microcolumns prior to their analyses by LC-FAIMS-MS/MS on a LTQ-Orbitrap Velos. Quantitative phosphoproteomics experiments were performed using CV stepping and peptide sequencing was obtained for separate LC-MS/MS analyses at fixed CV values. Peptide maps and MS/MS identification were clustered across sample sets to profile changes in peptide abundance. CK2-regulated candidates were further validated by *in vitro* kinase assay using synthetic peptides and reaction products were analyzed by LC-MS/MS on the LTQ-Orbitrap XL mass spectrometer.

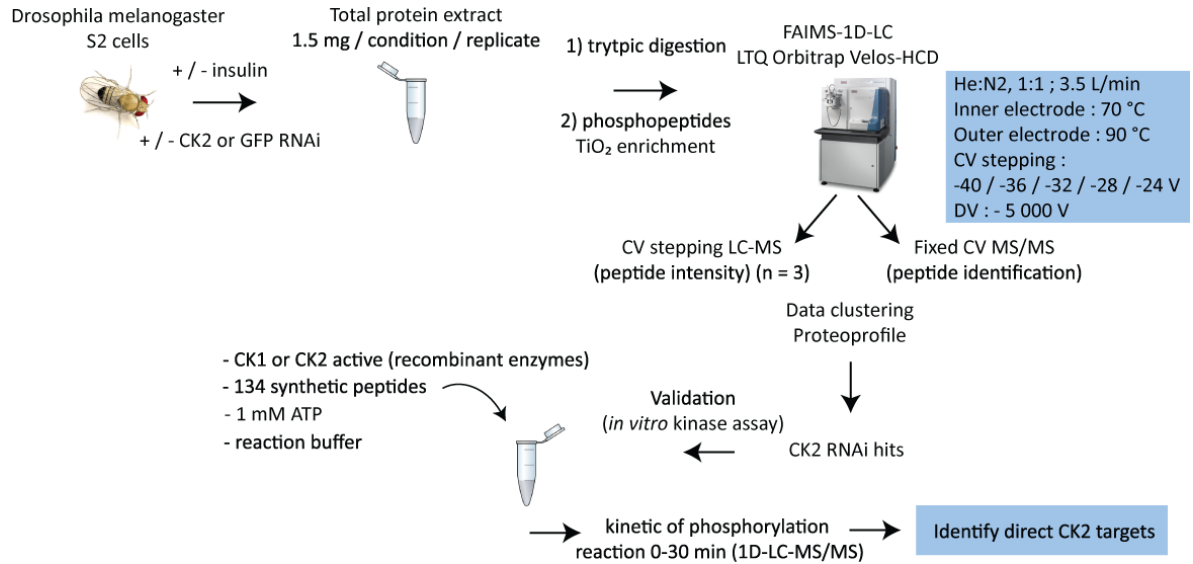


Figure 3.1. Workflow of the differential phosphoproteomics experiment.

Biological triplicates samples were prepared for each condition and detergent-free protein extraction was performed followed by tryptic digestion and TiO₂ phosphopeptides enrichment. Phosphopeptides were analyzed using LC-FAIMS-MS/MS on a LTQ-Orbitrap Velos using HCD fragmentation.

3.5.1. Insulin stimulation and CK2 RNAi treatment

Drosophila S2 cells were treated with 0.86 μ M of insulin with RNAi knock-down on CK2 or GFP (15 μ g/mL for 5 days). The effect of insulin on drosophila S2 cells were monitored by immunoblots for p-MAPK (T₂₀₂/Y₂₀₄) and p-AKT (S₄₇₃), to confirm their activation under the present conditions (**Figure 3.2. A**). Both phosphoproteins are rapidly phosphorylated following insulin treatment with a maximum activation at 8 min[1]. Inhibition of CK2-alpha expression was achieved using CK2 RNAi treatment and compared against a GFP RNAi negative control. Immunoblots indicated an efficient knock-down of CK2 with no apparent changes on p-MAPK or p-AKT (**Figure 3.2. A**). These results were also confirmed by RT-q-PCR where we observed 80% inhibition of CK2 mRNA levels (**Figure 3.2. B**). This level of inhibition is sufficient to obtain robust decrease in CK2 expression between control and CK2 RNAi treated cells for subsequent phosphoproteomics experiments.

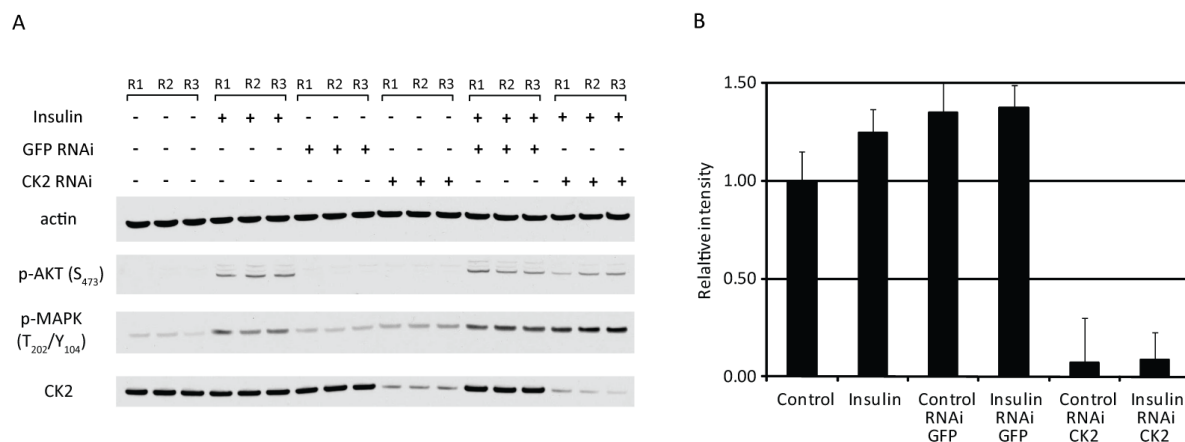


Figure 3.2. Phosphorylation of AKT and MAPK following insulin stimulation and CK2 knock-down after CK2 RNAi treatment.

A. Increases in phosphorylation of AKT (S₄₇₃) and MAPK (T₂₀₂/Y₂₀₄) are used as positive control for insulin stimulation. Western blot for CK2-alpha showed a significant inhibition upon CK2 RNAi knock down in S2 cells. B. RT-q-PCR results confirmed CK2 inhibition (80%) after CK2 RNAi treatment.

3.5.2. Combined phosphoproteomics and RNAi analyses to identify kinase substrates regulated by insulin stimulation

All samples were prepared in biological triplicates and analyzed by LC-FAIMS-MS/MS. These analyses enabled the identification of 2 067 distinct peptides from 692 proteins, of which 1 646 (574 unique phosphoproteins) were assigned as phosphopeptides with a FDR < 2% (**Table 3-I**). All identified phosphosites are listed in **Table A3- I**. The volcano plot distribution of phosphorylation fold-change vs. p-values following insulin treatment is presented in **Figure 3.3. A**, and indicated that a large proportion of phosphopeptides remained unaffected. The reproducibility of intensity measurements was determined from the relative standard deviation observed from biological triplicates. We observed that 95% of peptide intensities have an intensity fold change less than 2.7 between replicates (reproducibility graphs in **Figure A3. 1**).

Table 3-I. Table of statistics showing the number of identified phosphopeptides in *Drosophila melanogaster* S2 cells by LC-FAIMS-MS/MS on a LTQ-Orbitrap Velos using HCD fragmentation.

	Unique peptides	Unique phosphopeptides	Unique phosphosites	High confidence sites	Novel phosphosites	Unique proteins	Unique phosphoproteins
#	2067	1646	1569	1177	1397 (89%)	692	574

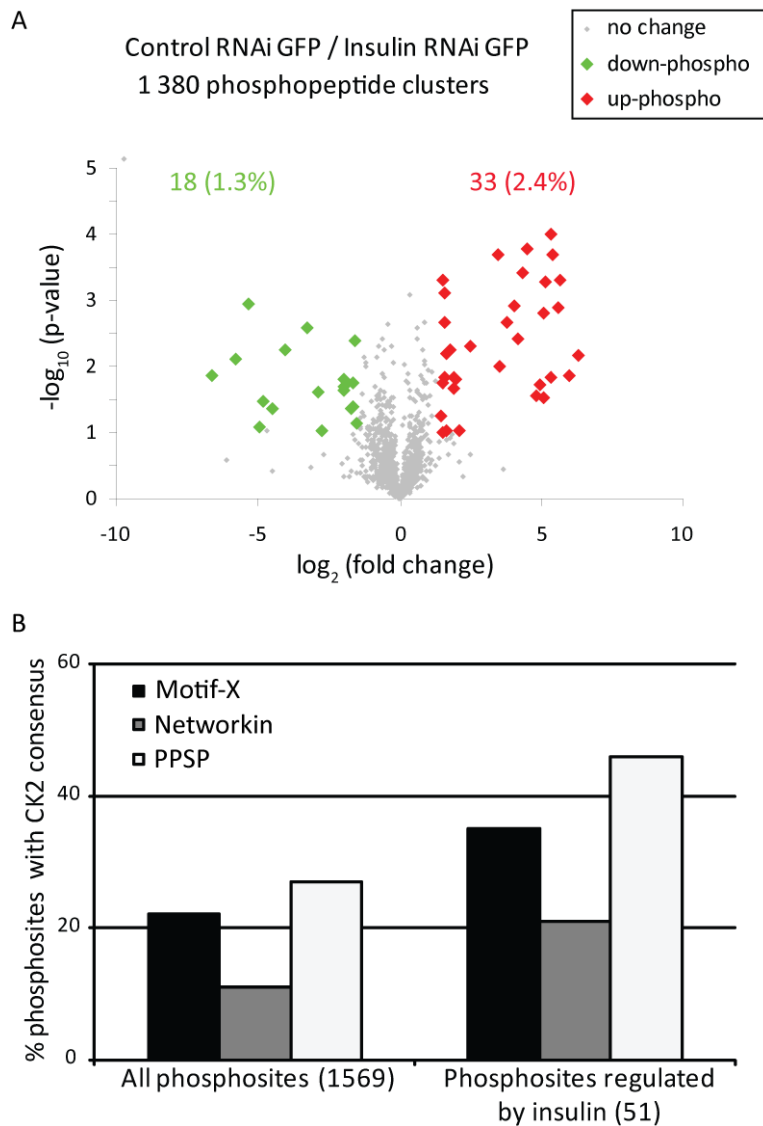


Figure 3.3. Representation of differentially phosphorylated peptides after insulin stimulation in S2 cells.

A. Volcano plot showing phosphopeptide clusters affected by insulin stimulation with a significant fold change (> 2.7) and a good t-test p-value (< 0.1). We observe that 18 and 33 phosphopeptide clusters are respectively down- and up-phosphorylated after insulin stimulation. B. Motifs analysis and kinase prediction softwares highlight the presence of CK2 consensus sites. Motif-X, NetworkKIN and PPSP tools were used to identify the kinases responsible for the phosphorylation of specific phosphosites. An increase in the proportion of phosphosites potentially phosphorylated by CK2 is observed when comparing the whole dataset (22%, 11%, and 27%, using respectively Motif-X, NetworkKIN and PPSP) to phosphosites affected by insulin stimulation (35%, 21% and 46%).

A phosphopeptide is considered differentially expressed when showing a t-test p-value < 0.1 calculated on three biological replicates and an average intensity fold change higher than 2.7 (7 standard deviations on sample replicates). The comparison of Control GFP RNAi and Insulin GFP RNAi treated samples (1 380 phosphopeptide clusters) enabled the identification of 18 (1.3%) phosphopeptides down-phosphorylated after insulin treatment (green) whereas 33 (2.7%) phosphopeptides were found to be up-phosphorylated (in red). In total, 51 distinct phosphopeptides were differentially regulated following 8 min incubation with insulin. Positive controls such as Pi3k92E (Y₁₃₈) and insulin receptor substrate 1 (Chico) (S₅₀₂) were significantly up-phosphorylated after insulin treatment (**Table 3-II**). Pi3k92E is a well know interactor activated by Ras through tyrosine kinase receptor signalling[354], and showed more than 80-fold increase in phosphorylation upon insulin stimulation. We also identified the S₅₅₅ from Chico with a four-fold increase in phosphorylation consistent with previous studies[355]. This protein exerts an important role in the control of cell size and growth and in the regulation of lipid levels[60]. Both Chico and Pi3k92E are known to interact together to mediate insulin signaling events[356]. We also observed the protein split ends (Spen) up-phosphorylated on a novel site (S₂₆₁₆). This protein is involved in the EGF signalling pathway through Ras[357] and can act as a transcriptional co-repressor[358]. To determine if changes in phosphorylation can be regrouped into specific recognition motifs, we compared all phosphosites with different kinase prediction software such as Motif-X, Networkin and PPSP (**Figure 3.3. B**). These analyses revealed that approximately 20% of all identified phosphopeptides harboured a CK2 consensus motif. Interestingly, more than 30% of the 51 phosphosites regulated by insulin were potential CK2 targets. The CK2 motif was the only one displaying a significant enrichment level upon insulin treatment, though other motifs such as ERK1/2, GSK-3, CDK5 were not when all identified phosphopeptides were considered. Previous investigations indicated that nuclear MAPK and CK2 showed a 2-3 fold increase activity within 10 min after insulin stimulation[343].

Table 3-II. Examples of phosphosites affected by insulin or CK2 RNAi treatment with their corresponding intensity fold change and p-value calculated across 3 biological triplicates.

Uniprot ID, gene name, peptide sequence and phosphosites position are displayed. CK2-related phosphorylation for *in vitro* kinase assays using synthetic peptides is indicated. High: transformation into a majority phosphorylated form (80-100%) after 30 min of reaction, medium: detection of a phosphorylated peptide (< 50% of the total peptide intensity), no: no phosphorylation detected, N/A: non-tested peptide sequence. The presence of CK1/CK2 consensus motifs are indicated (* represents CK2 known targets according to NetworKIN) as well as phosphosites conservation in mammals. Pink and green backgrounds indicate phosphopeptides up- and down-regulated after insulin stimulation, respectively. Pink and green writings indicate phosphopeptides up- and down-regulated after CK2 RNAi treatment, respectively.

Control RNAi GFP / Insulin RNAi GFP	t-test p-value	Insulin RNAi GFP / Insulin RNAi CK2	t-test p-value	UniProt ID	Gene Name	Peptide Sequence	CK2 reactivity IVKA	Phosphosite position	CK2/CK1 consensus	Phosphosite conserved in mammals
-1.72	0.236	-3.45	0.001	Q02645	hts	NPFDNVpTDELNEYKR	high	T586	CK2/CK1	yes
-1.65	0.222	-4.56	0.013	Q86P47	Cnx99A	KNQKpSDLDDNEEEK	medium	S566	CK2/CK1	yes
-1.59	0.031	-133.48	0.002	Q7PLS7	Integrator complex subunit 3 homolog	VHNIEDNpTDDDDDLPLSEVR	high	T561	CK2	yes
-1.15	0.627	-3.09	0.005	A4V3G5	scrib	DLEpSDSEPFEEPEPSR	high	S433	CK2/CK1	yes
-1.11	0.636	-3.71	0.006	Q9V426	vig	KKENDpSEEELEYDPALYPQR	high	S342	CK2*/CK1	yes
-1.08	0.781	-11.13	0.028	Q8MRY4	SD13619p	SLEpSDEDDILEK	high	S198	CK2/CK1	no
-1.07	0.873	-42.41	0.019	Q9VQK7	LD45152p	AHGpSEDEQVAEK	medium	S176	CK2*/CK1	yes
-1.02	0.946	-6.58	0.011	Q9VBX3	vig2	KKESNpSEDELEYDPSLYPQR	high	S295	CK2/CK1	yes
-1.02	0.907	-4.36	0.002	A1Z8W9	garz	GYTpSDSEITK	high	S1276	CK2/CK1	yes
-1.01	0.972	-70.84	0.005	Q8MSS1	lva	ELQQLQQDQDpSEVEHVR	medium	S1284	CK2*/CK1	yes

Table 3-II (suite)

Control RNAi GFP / Insulin RNAi GFP	t-test p-value	Insulin RNAi GFP / Insulin RNAi CK2	t-test p-value	UniProt ID	Gene Name	Peptide Sequence	CK2 reactivity IVKA	Phosphosite position	CK2/CK1 consensus	Phosphosite conserved in mammals
13.92	0.002	-12.26	0.002	P05205	Su(var)205	IDNPESAKVpSDAEeeeeeyAVEK	high	S15	CK2*/CK1	yes
24.40	0.000	-24.40	0.000	Q8IQ60	scny	AKpSPVKIQVK	no	S650	CK1	yes
28.33	0.028	-28.33	0.028	Q8SZ44	Protein CDV3 homolog	VPSAPDGGNYNEDEDEpSNGYDNADVNK	no	S134	CK2/CK1	no
30.53	0.019	-30.53	0.019	P22058	D1	SSTAAVAAIpSPGIKK	no	S30	CK1	no
34.49	0.002	-34.49	0.002	Q95RC8	mu2	SEpSPPETPIRK	no	S630	CK2/CK1	no
34.49	0.010	-38.49	0.008	Q4V5E2	CG30085	IYNLSpSPPDRK	no	S959	CK2/CK1	yes
83.74	0.000	-1.52	0.000	P91634	Pi3K92E	TDTTSFSSDpYELTK	no	Y138	no	no
3.91	0.016	1.43	0.175	Q9XTN2	chico	AYpSVGSK	no	S502	CK1	yes

Next, we compared changes in protein phosphorylation in insulin-treated drosophila cells following RNAi knock-down of GFP and CK2. These analyses enabled the identification of 1 417 phosphopeptides clusters of which 65 (4.6%) were up-regulated (in red) and 83 (5.8%) down-regulated (in green) after CK2 RNAi treatment (**Figure 3.4. A**). Interestingly, 29 of the 83 phosphosites showing a decrease in phosphorylation, including Eif4b (S₄₆) and Top2a (S₁₃₄₄) previously reported to be phosphorylated by CK2[359-363], were also identified as putative substrates by NetworkKIN (**Figure 3.4. B**). Other substrates showing a decrease in phosphorylation upon RNAi treatment but without a CK2 consensus motif could be indirect targets recognized by different kinases.

We also identified new phosphosites such as the protein Hu-li tai shao (T₅₈₆, S₆₃₀ and T₆₃₄) involved in actin assembly and protein Lava lamp (S₁₂₈₄) implicated in microtubule interaction with Golgi apparatus (**Table 3-II**) that were negatively regulated by CK2 RNAi. The protein Hu-li tai shao (Hts) is the homologue of the mammalian Adducin[364]. **Figure 3.5. A** presents the extracted ion chromatograms for the tryptic peptide ⁵⁸⁰NPFDNVpTDELNEYKR⁵⁹⁵ at *m/z*: 683.62³⁺ from Hts for Control RNAi GFP, Insulin RNAi GFP and Insulin CK2 RNAi conditions. The decrease in abundance observed upon CK2 RNAi support its assignment as a potential CK2 target. The HCD MS/MS spectrum of the corresponding precursor ion is shown in **Figure 3.5. B** and confirmed the location of the phosphorylated residue T₅₈₆.

Interestingly, we identified 6 phosphosites from CG30085 (S₉₅₉), Su(var)205 (S₁₅), scny (S₆₅₀), protein CDV3 homolog (S₁₃₄), D1 (S₃₀) and mu2 (S₆₃₀) that showed an increase in phosphorylation upon insulin stimulation and a decrease after CK2 RNAi treatment (**Table 3-II**). Four out of these sites (Su(var)205, mu2, protein CDV3 homolog and CG30085) were predicted to have a CK2 consensus motif based on PPSP. Moreover, heterochromatin protein 1 (Su(var)205) (S₁₅) is a well known CK2 substrate in drosophila[119]. The change in abundance of the corresponding phosphopeptide ⁵IDNPESSAKVpSDAEEEEEEYAVEK²⁸ is presented in **Figure 3.5. C** for Control, Insulin GFP RNAi and Insulin CK2 RNAi conditions. Confirmation of the identity of this phosphopeptide and the phosphorylated S₁₅ residue is shown in **Figure 3.5. D** for the HCD MS/MS spectrum of *m/z*: 926.73³⁺.

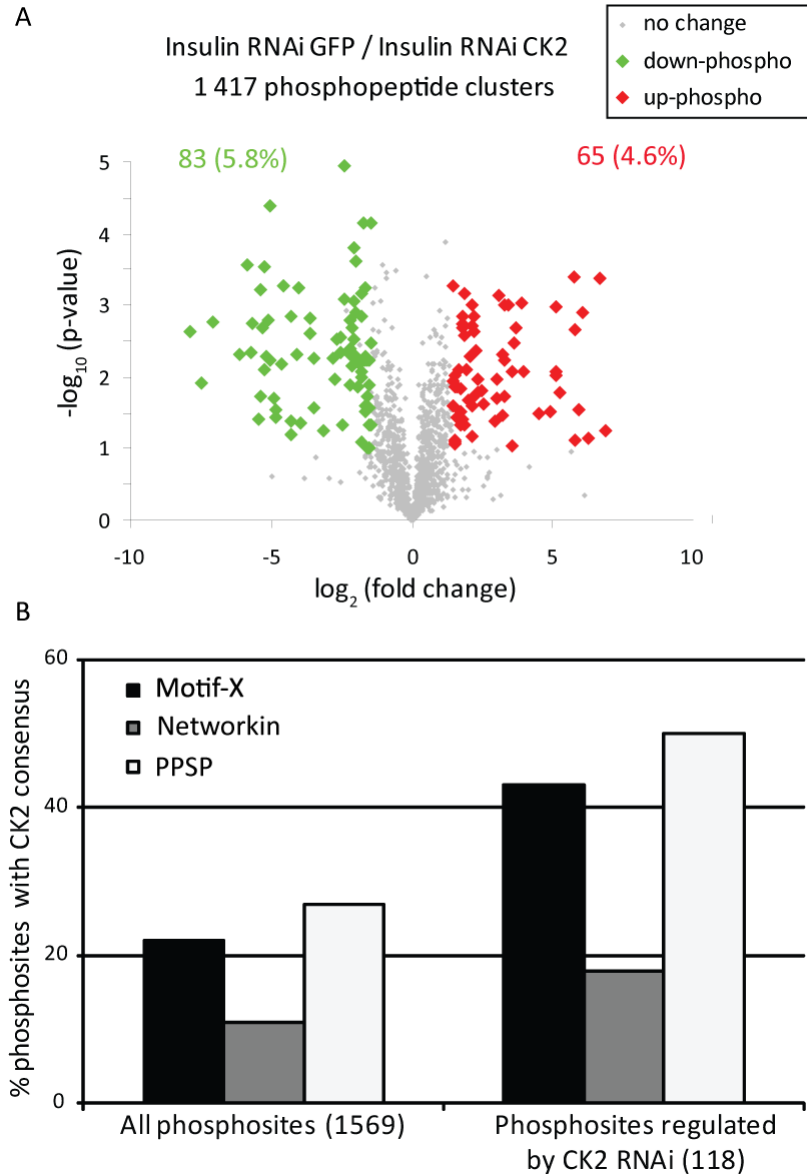


Figure 3.4. Representation of differentially phosphorylated peptides after CK2 RNAi treatment in insulin-stimulated S2 cells.

A. Volcano plot showing phosphopeptide clusters affected by CK2 RNAi treatment in insulin stimulated cells with a significant fold change (> 2.7) and a good t-test p-value (< 0.1). We observe that 83 and 65 phosphopeptide clusters are respectively down- and up-phosphorylated after CK2 RNAi treatment. B. Motifs analysis and kinase prediction softwares highlight the presence of CK2 consensus sites. An increase in the proportion of phosphosites potentially phosphorylated by CK2 is observed when comparing the whole dataset (22%, 11%, and 27%, using respectively Motif-X, NetworkKIN and PPSP) to phosphosites affected by CK2 RNAi treatment (43%, 18% and 50%).

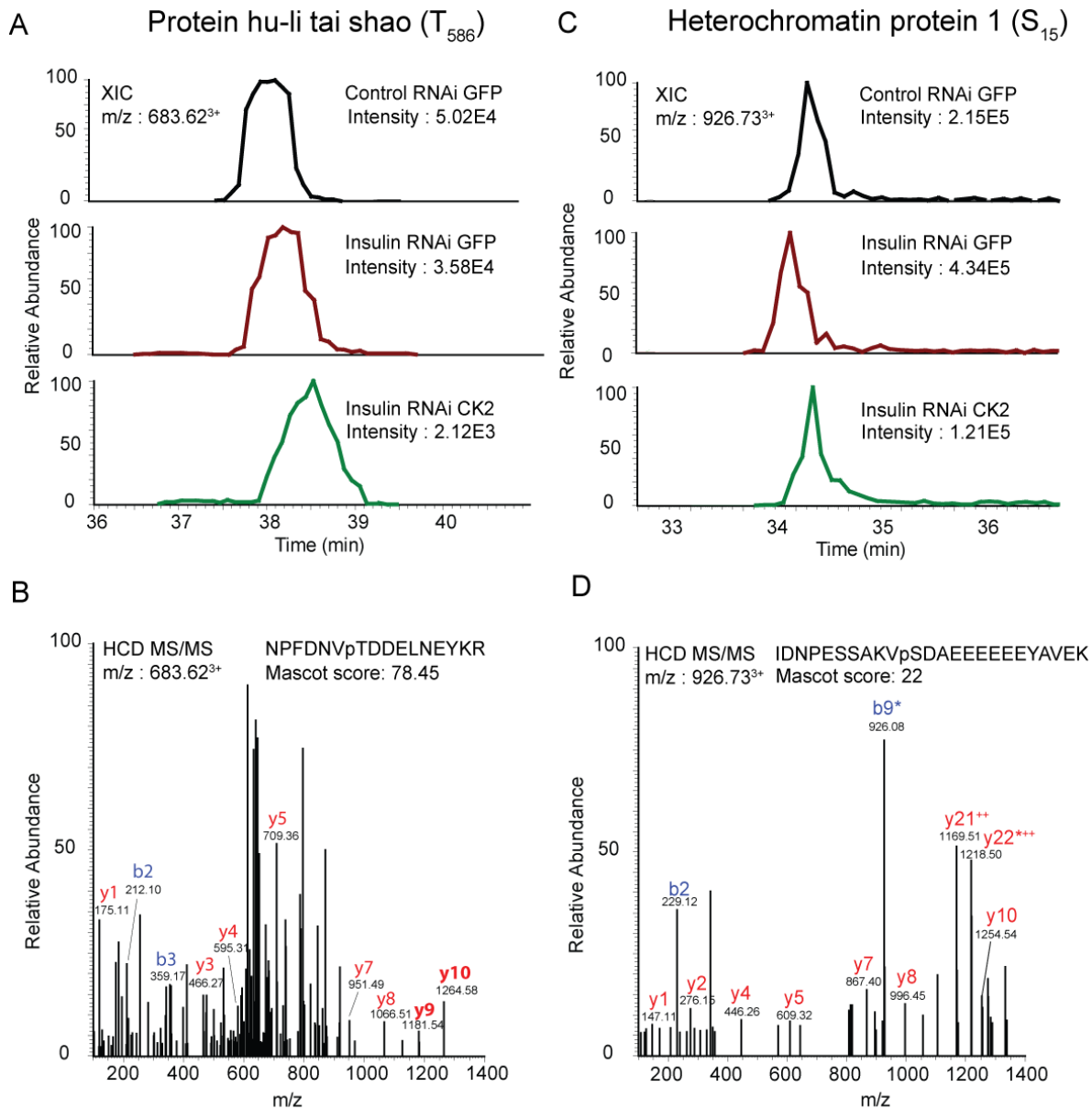


Figure 3.5. Examples of phosphoproteins modulated by insulin and/or CK2 RNAi treatment.

A. XIC of ion m/z : 683.62³⁺ from the protein Hu-li tai shao (T₅₈₆) which shows no change in intensity after insulin stimulation but is down phosphorylated after CK2 RNAi treatment. B. HCD MS/MS spectrum confirms the peptide sequence of ion m/z : 683.62³⁺ NPFQNVpTDELNEYKR and phosphate localization. C. XIC of ion m/z : 926.73³⁺ from the protein heterochromatin protein 1 (S₁₅) up-phosphorylated after insulin stimulation and down phosphorylated after CK2 RNAi treatment. D. HCD MS/MS spectrum confirms the peptide sequence of ion m/z : 926.73³⁺ IDNPESAKVpSDAEEEEEEYAVEK and phosphate localization.

Heterochromatin protein 1 is a structural component of heterochromatin and is also involved in the formation of the kinetochore. Chromosomal protein D1 (S₃₀) is involved in chromatin organization[365] and ubiquitin carboxyl-terminal hydrolase 36 (scny) (S₆₅₀) is a histone deubiquitinase playing a role of repressor of transcription[366]. RE66338 (CG30085) (S₉₅₉) has sequence homology with the N-terminus of the human telomere-associated protein RIF1 involved in DNA damage response and cell cycle checkpoints regulation. Protein CDV3 homolog (S₁₃₄) is involved in alternative splicing and mu2 (mutator 2) (S₆₃₀) shares homology with the human MDC1 protein in the C-terminus implicated in DNA repair response. These findings suggest that insulin treatment could regulate CK2 activity leading to the phosphorylation of protein substrates involved in chromatin organization and DNA damage response. Previous reports suggested an interplay between insulin signaling and systemic responses to localized DNA damage to coordinate growth and metabolic activities in drosophila larva[367]. Also, CK2 is known to be involved in reactive oxygen species and DNA double or single strand breaks formation[368-370]. Surprisingly, we found that the protein Nopp140-like nucleolar protein (Nopp140) was modulated by CK2 RNAi and by insulin. In Nopp140, Ser residues 419, 448, 450 and 560 are significantly up-phosphorylated after CK2 RNAi, S₅₆₀ and S₄₁₉ are respectively down- and up-phosphorylated after insulin treatment. Specific interactions between CK2 and Nopp140 have already been described and it was found that Nopp140 binds primarily to the β -CK2 regulatory subunit[371]. These results highlight a dynamic relation between the insulin pathway and CK2-dependent phosphorylation. The complete list of phosphopeptides affected by insulin or CK2 RNAi is presented in **Table A3- II**. For convenience we reported fold changes, p-values, peptide sequences, protein identification, phosphosite conservation and CK1/CK2 consensus motifs.

3.5.3. *In vitro* kinase assays confirmed CK2 direct substrates

Our combined quantitative phosphoproteomics and RNAi experiments suggested several potential CK2 substrates of which at least six were also regulated by insulin. To confirm if

these potential CK2 targets were direct substrates, we conducted *in vitro* kinase assays on 134 synthetic peptides using active recombinant CK2. These synthetic peptides were selected based on identification from phosphopeptides regulated after insulin stimulation, CK2 RNAi treatment or non-regulated. The enzyme CK1 was used as a negative control, since CK1 and CK2 have consensus motifs that share similar acidic environment. All peptides were mixed at equal concentration in a single tube and the temporal changes in phosphorylation were monitored over the course of 30 min to identify best CK2 candidates. We also included CK1 and CK2 substrate peptides as positive and negative controls. **Figure 3.6. A** shows the relative phosphopeptide abundance normalised to the total peptide intensity (phosphorylated + non-phosphorylated form) for 11 synthetic peptides of interest (29 phosphorylated peptides in total) and two controls. We observed a rapid and complete conversion of the CK2 peptide substrate (CK2SP) after 30 min of incubation with CK2 enzyme, while no detectable phosphorylation of CK1 peptide substrate (CK1SP) was noted over the same time period. **Figure 3.6. A** also identifies two groups of CK2 substrates based on their respective rates of phosphorylation. We identified 18 out of 29 peptides that showed 80% to 100% of substrate phosphorylation (highly reactive hits) and 11 peptides with 10 to 50% conversion (medium reactivity). It is noteworthy that all 29 peptides sequences have a CK2 consensus site. **Figure A3. 2** shows a similar representation for *in vitro* kinase assay using CK1 enzyme where only CK1SP is phosphorylated and confirmed the specificity of substrates for CK2. Moreover, 15 out of 29 CK2 targets were also correlated in differential phosphoproteomics experiment using CK2 RNAi (**Figure 3.4. A**). The fold change in abundance for 11 out of 15 CK2 targets identified in the *in vitro* kinase assay are presented in **Figure 3.6. B** for control or insulin-stimulated S2 cells after CK2 RNAi treatment. Interestingly, we confirmed known CK2 targets such as RpLP1 (S₁₀₂), CG1703 (S₄₆), Vig (S₃₄₂), Vig2 (S₂₉₅) and Su(var)205 (S₁₅). Su(var)205 (heterochromatin protein 1 [S₁₅]) showed an increase in phosphorylation after insulin stimulation and a decrease after CK2 RNAi knock-down. The use of the *in vitro* kinase assay supported a direct phosphorylation of substrates by CK2. We also confirmed insulin-independent CK2 targets such as the proteins Hu-li tai-shao (S₅₈₆) (**Figure 3.5. A**), Lava lamp (S₁₂₈₄) and Cnx99a (S₅₆₆).

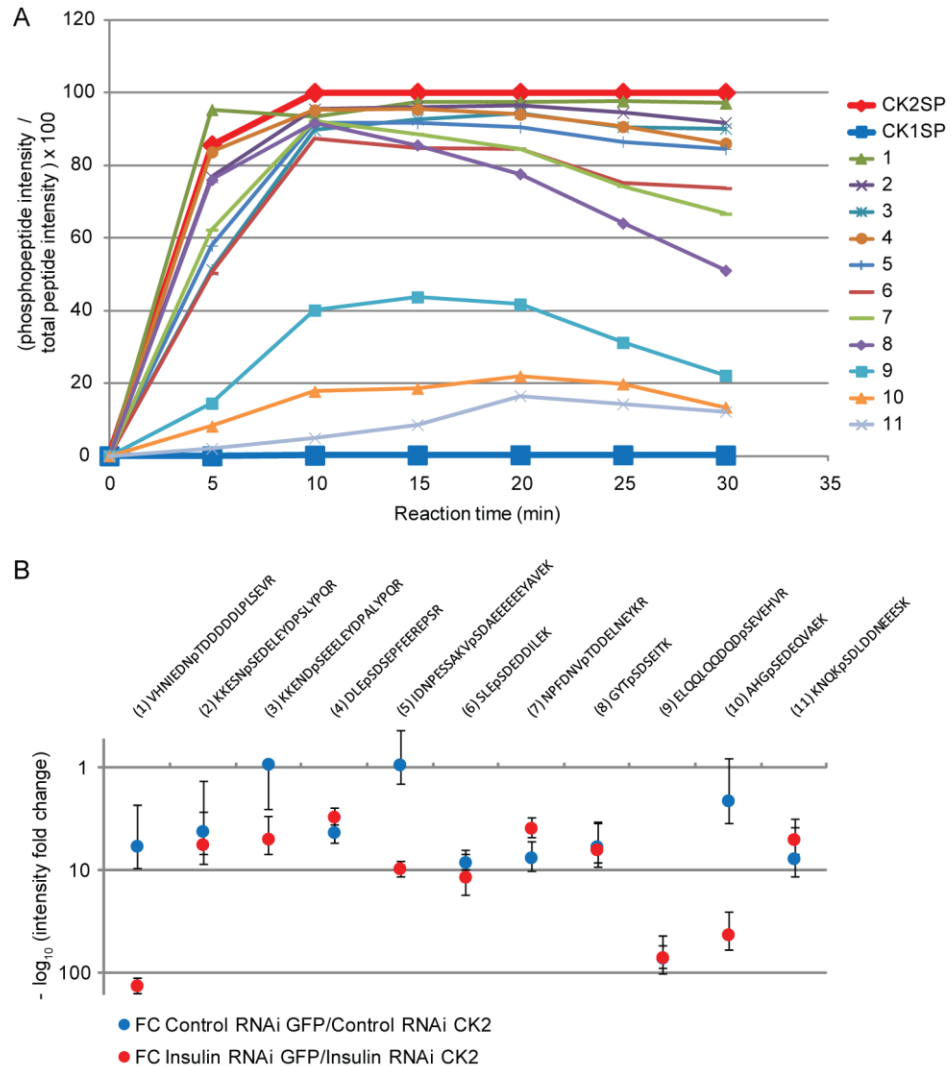


Figure 3.6. Phosphorylation kinetics of synthetic peptides using *in vitro* kinase assay with CK2 recombinant enzyme.

A set of 11 CK2 peptides targets (plus one positive control, CK2SP and one negative control CK1SP) is represented. A. Representation of the ratio of intensity between the phosphorylated form of a specific peptide and the total peptide intensity (phosphorylated + non phosphorylated forms). As expected, CK2SP is completely transformed into a phosphorylated form and CK1SP does not show any phosphorylation. Numbers indicate unique peptide sequences detailed in the next panel. B. Fold change values of specific phosphopeptide intensities after CK2 RNAi treatment in control or insulin stimulated cells for 11/15 specific hits correlating with *in vitro* kinase assay results. These 11 peptides hits show a significant decrease in phosphorylation after CK2 RNAi and a significant increase in phosphorylation is observed using CK2 *in vitro* kinase assay (**Table 3-II**).

Additional information on the changes of phosphorylation of some of these substrates is presented in **Figure A3. 3** for CK2 and CK1 *in vitro* reaction products. Peptide incubated with CK1 displayed strikingly different trends supporting their CK2 substrate specificity.

We also observed multiple phosphorylated residues for 9 out of 29 peptides. We could not determine any consistent trend in the rate of changes for these sites to identify an orderly sequence of phosphorylation or site-specific modification between mono-phosphorylated isomers. Examples of kinetic profiles are shown in **Figure A3. 4**. We identified both mono- and di-phosphorylated forms of KKEEESDQSDDDMGFGLFD and SSSSESDSDEEEEAPKKPAAK as early as 5 min after incubation with CK2 with 100% conversion. The mono-phosphorylated KKEEESDQpSDDDMGFGLFD is only detectable during the first five min and is rapidly converted into the di-phosphorylated form. It is noteworthy that the localization of phosphorylation sites position can be ambiguous for peptides with multiple phosphorylatable residues in close proximity.

Intriguingly, DNA topoisomerase 2 (AKKpSDDDDVEEVTPR) (S₁₃₄₄) showed a significant decrease in phosphorylation after CK2 RNAi. This site corresponds to a CK2 consensus motif but was not detected in any phosphorylated form when the synthetic peptide was incubated with recombinant CK2. However, this site was previously described to be phosphorylated and regulated by CK2[111]. This result could be explained by a specific fold that could be present in the protein substrate that is obviously different for the linear peptide. It is noteworthy that the C-terminus of topoisomerase II is phosphorylated and can be extended or undergo conformational changes to interact with other proteins. It is possible that acceptor phosphosites can be arranged in a hierarchal phosphorylation cascade (successive phosphorylation on adjacent sites by CK2)[372, 373], a situation that cannot occur *in vitro*.

A summary of results obtained for the phosphoproteomics experiments along with CK2 *in vitro* kinase assay is presented in **Figure 3.7**. The Venn diagram indicates the overlap between insulin- and CK2 RNAi-dependent phosphosites and for CK2 direct targets obtained from the *in vitro* kinase assay. Altogether, we identified 51 (24 CK2 consensus sites) and 118 unique phosphosites (115 CK2 consensus sites) regulated by

insulin and CK2 RNAi treatment, respectively. We observed an overlap of 6 phosphoproteins (4 CK2 consensus sites) positively regulated after insulin and negatively modulated after CK2 RNAi. Using CK2 *in vitro* kinase assay, we identified 29 phosphopeptides (all with CK2 consensus site) as direct CK2 targets, 15 of which also correlated with CK2 RNAi experiments.

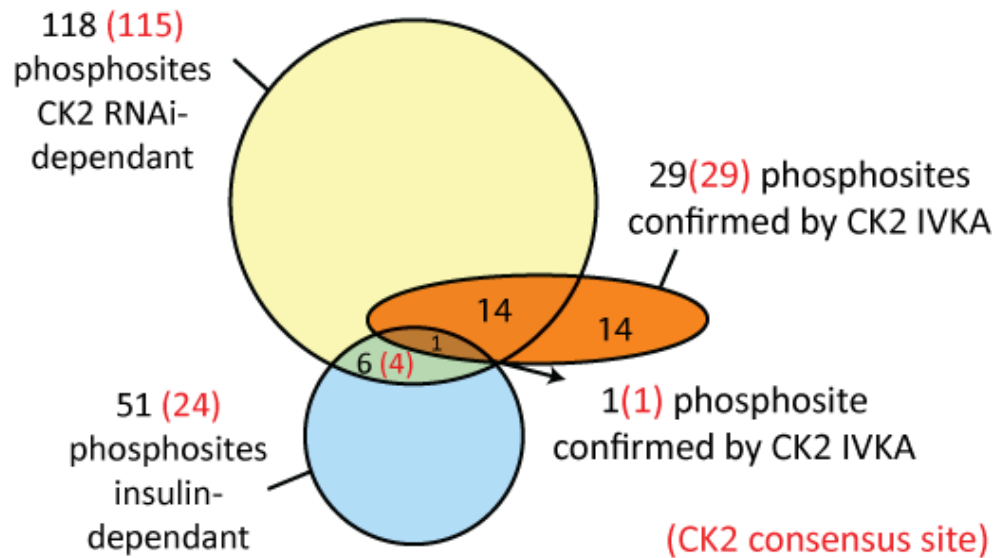


Figure 3.7. Venn diagram representation summarizing the results of the global phosphoproteomics experiment conducted in S2 cells.

We show that 51 and 118 unique phosphosites are modulated after insulin stimulation and CK2 RNAi treatment, respectively. Six phosphosites are up-phosphorylated after insulin stimulation and down-phosphorylated after CK2 RNAi treatment and one phosphosite out of these 6 (heterochromatin protein 1, S₁₅) was confirmed by CK2 *in vitro* kinase assay. A total of 29 phosphosites down-phosphorylated after CK2 RNAi have been confirmed by CK2 *in vitro* kinase assay to be CK2 direct targets.

In vitro kinase assays were performed to evaluate CK2 activity following insulin stimulation in S2 cells. CK2SP was spiked in a total protein lysate and temporal changes were monitored for biological triplicates by determining the proportion of phosphorylated CK2SP to total CK2SP (phosphorylated + non phosphorylated) as shown in **Figure A3. 5**. This figure indicates that even when TiO₂ enrichment is performed, we still detect the non-phosphorylated form of CK2SP due to the large amount of synthetic peptide compared to the endogenous CK2 kinase. Results obtained for these experiments suggest that CK2 is constitutively active in S2 control cells as progressive change in CK2SP is observed over the reaction period (**Figure A3. 5**). These experiments also indicated that 45% and 30% of CK2SP is phosphorylated after 60 min in control and insulin stimulated cells, respectively. Also, insulin-stimulated S2 cells displayed a reduction in CK2SP phosphorylation, suggesting that CK2 activity remains unchanged upon insulin treatment. This result is also consistent with that observed for GFP RNAi negative control cells. S2 cells treated with CK2 RNAi show a significant decrease in phosphorylation of CK2SP compared to GFP RNAi treated cells, with only 20% of CK2SP detected in a phosphorylated form. These results suggest that increased phosphorylation of CK2 substrates cannot be attributed to changes in CK2 activity upon insulin stimulation, and that other mechanisms including conformational changes of substrates to expose CK2 consensus sites might be operative.

3.6. Conclusion

Quantitative phosphoproteomics experiments performed on *Drosophila melanogaster* S2 cells highlighted new insulin-dependent phosphorylation events. Using a combination of kinase predictions software, we identified that a high proportion of insulin-regulated phosphorylation events were associated to residues sharing a CK2 consensus motif. We used a CK2 RNAi strategy to determine insulin- and CK2-dependent phosphosites. These analyses revealed that 51 phosphopeptides were regulated by insulin and 148 were directly affected by CK2 RNAi knock-down. We observed that the phosphorylation of 6 proteins was modulated by both insulin stimulation and CK2 RNAi inhibition. We also identified

that 15 out of 118 unique phosphosites regulated by CK2 RNAi were confirmed as *bona fide* CK2 sites by *in vitro* kinase assay. We confirmed that Ser15 residue from heterochromatin protein 1 was phosphorylated by CK2 upon insulin treatment. A kinase assay performed on control and insulin-stimulated S2 cells extracts indicated that CK2 is constitutively active and that its activity remains unchanged upon insulin treatment. We surmised that conformational changes in response to insulin treatment might render substrates more susceptible to phosphorylation by constitutively active CK2. Alternatively, specific phosphatase sharing CK2 consensus sites might be inactivated upon insulin treatment. This balance between CK2 and phosphatase activity might determine the phosphorylation levels of specific protein targets. For example, protein phosphatase 1 (PP1) could be involved in a coordinated interplay with CK2 where they could act in concert to regulate proteins activity[374]. A change in CK2 cellular localization might also regulate its local activity and affect the phosphorylation of specific targets.

3.7. Acknowledgements

IRIC receives infrastructure support funds from the Fonds de la Recherche en Santé du Québec (FRSQ). This work was carried out with the financial support of operating grants from the National Science and Engineering Research Council (NSERC) and the Canada Research Chair programs. The authors also thank the following colleagues from Thermo Juergen Srega and Brenda Kesler for insightful suggestions and logistic support, and Michael Belford for technical assistance.

Chapitre 4 : Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments

Mathieu Courcelles^{1,2,5}, Gaëlle Bridon^{1,2,5}, Sébastien Lemieux⁴, Pierre Thibault^{1,2,3}

Article publié dans *Journal of Proteome Research*, 2012

J Proteome Res. 2012; 11(7): 3753-65

[2]

Inclus avec la permission de John Wiley and Sons

¹IRIC, Institut de recherche en immunologie et oncologie, ²Département de Biochimie, ³Département de Chimie, ⁴Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada.

⁵These authors have contributed equally to this work.

4.1. Contribution des auteurs

Gaëlle Bridon, Mathieu Courcelles et Pierre Thibault ont écrit le manuscrit. Sébastien Lemieux a révisé le manuscrit. **Gaëlle Bridon** a effectué la culture cellulaire et la préparation des échantillons. **Gaëlle Bridon** et Mathieu Courcelles ont effectué les expériences par spectrométrie de masse et l'analyse des données. Mathieu Courcelles a développé les algorithmes pour détecter les isomères positionnels des phosphopeptides. Pierre Thibault et Sébastien Lemieux ont supervisé cette analyse.

4.2. Abstract

The past decade has been marked by the emergence of selective affinity media and sensitive mass spectrometry instrumentation that facilitated large-scale phosphoproteome analyses and expanded the repertoire of protein phosphorylation. In spite of these remarkable advances, the precise location of phosphorylation site still represents a sizable challenge in view of the labile nature of the phosphoester bond and the presence of neighboring phosphorylatable residues within the same peptide. This difficulty is exacerbated by the combinatorial distribution of phosphorylated residues giving rise to different phosphopeptide isomers. These peptides have similar physicochemical properties and their separation by LC is often problematic. Few studies have described the frequency and distribution of phosphoisomers in large-scale phosphoproteomics experiments and no convenient informatics tool currently exist to facilitate their detection. To address this analytical challenge, we developed two algorithms to detect separated and co-eluting phosphopeptide isomers and target their subsequent identification using inclusion list in LC-MS/MS experiments. Using these algorithms, we determined that the proportion of isomers present in phosphoproteomics studies from mouse, rat and fly cell extracts represent 3-6% of all identified phosphopeptides. While conventional analysis can identify chromatographically separated phosphopeptides, targeted LC-MS/MS analyses using inclusion lists provided complementary identification, and expanded the number of phosphopeptide isomers by at least 52%. Interestingly, these analyses revealed that the occurrence of phosphopeptides isomers can also correlate with the presence of extended phosphorylatable amino acids that can act as “phosphorylation switch” to bind complementary domains such as those present in SR proteins and ribonucleoprotein complexes.

4.3. Introduction

Protein phosphorylation is a reversible protein post-translational modification that imparts structural changes on protein substrates to modulate their activities and mediate interactions with other binding partners. This modification regulates different biological processes such as cell division, growth, and differentiation, and its misregulation is often associated with many human diseases, including cancer. In eukaryotes, the large majority of protein phosphorylation takes place on hydroxylated amino acids such as serine, threonine, and tyrosine with a relative distribution of 88:11:1[19, 161].

The development of phosphopeptide enrichment procedures along with improvements in mass spectrometer sensitivity and bioinformatics algorithms have facilitated the identification of phosphopeptides from complex cell extracts, and literally revolutionized the fields of cell biology and cell signaling[375]. Phosphoproteomics is now routinely performed in many labs as part of discovery studies, and several reports have documented the capability to identify and profile the abundance of thousands of phosphopeptides in response to different cell stimulation[19, 161, 376, 377]. While these technological advances have significantly expanded the inventory of protein phosphorylation from different species, our abilities to obtain comprehensive phosphoproteome coverage and unambiguous site localization still represent a sizable analytical challenge. For example, approximately one third of phosphorylation sites identified using current mass spectrometry and bioinformatics approaches have ambiguous assignments due to the labile nature of the phosphoester bond and the poor quality of phosphopeptide fragmentation[19, 199]. The prompt loss of phosphoric acid (H_3PO_4) from collisional induced dissociation (CID) of multiply-charged phosphopeptides often dominates the product ion spectrum and results in relatively few site-specific fragment ions[375]. This situation is partly alleviated using electron transfer dissociation (ETD) whereby the phosphoester bond remains intact and the modified residues can be assigned from the consecutive cleavages of backbone C-N bonds[246]. Different algorithms were designed to evaluate the probability of phosphorylation site assignment based on database search engine results. These include Ascore[272], PhosphoScore[275], PhosphoScan[273],

PhosCalc[276] for MS/MS spectra obtained by CID, and, SLoMo[277], Phosphinator[278], PhosphoRS[279] for those acquired using both CID and ETD fragmentation modes.

The correct assignment of phosphorylation sites can also be compromised by the presence of co-eluting phosphopeptide isomers where different residues from the same peptide can be alternatively modified (**Figure 4.1**). In this situation, the MS/MS spectrum corresponds to a mixed population of precursor ions giving rise to confounding fragment ions and ambiguous site assignment. None of the algorithms mentioned above were designed to distinguish positional isomers and may not use sufficient spectral information such as peak intensities and delta m/z windows to account for different possibilities. The probabilities calculated by these algorithms favor candidates with the highest number of correlated fragment ions or report ambiguous localization when observed. Unfortunately, they do not report if ambiguous localization is due to lack of structurally relevant fragment ions or conflicting assignments originating from different isomers.

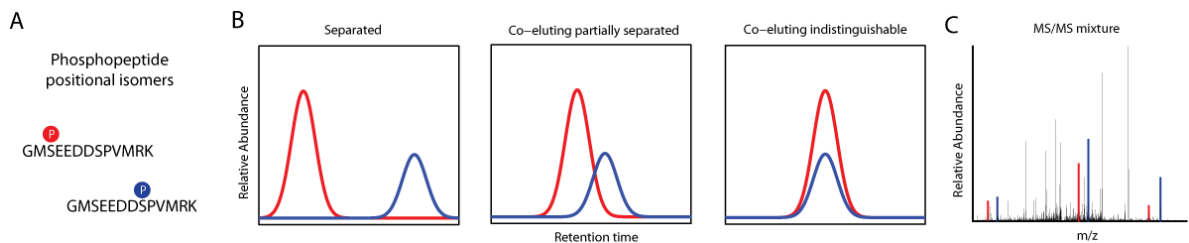


Figure 4.1. Separation and detection of phosphopeptide isomers.

A. Phosphopeptide isomers are peptides with the same amino acid sequence but phosphorylated at different positions. B. Phosphopeptide isomers can be detected if they are completely or partially separated by liquid chromatography. C. A mixture of two co-eluting phosphopeptide isomers can be identified by MS/MS from distinct sets of fragment ions.

The separation of phosphopeptide isomers is challenging as they share similar physiochemical properties. The separation of synthetic phosphopeptide isomers were first reported for NEFM and microtubule-associated protein τ using reverse-phase liquid chromatography (LC)[378, 379]. Other examples of phosphopeptide isomers were observed in large scale phosphoproteomics experiments, and included proteins such as Fus3p[380], Erk2[381], Egfr[382], Dok1[383], and Sprouty2[384]. Different chromatography media such as monolithic column[385] and HILIC with aminopropyl stationary phase[386] were proposed to separate phosphopeptide isomers. More recently, capillary zone electrophoresis (CZE)[387] and FAIMS[388, 389] were used as alternative techniques to separate phosphopeptide isomers. While these examples account for a limited number of phosphopeptide isomers from specific substrates, the frequency and challenges associated with their observations in large-scale phosphoproteomics experiments still remain largely unknown. To provide further insights on the occurrence and distribution of phosphopeptide isomers in large-scale phosphoproteomics experiments, we developed two algorithms to detect their presence from partly resolved LC peaks and identify co-eluting isomers from mixed MS/MS spectra. Candidates detected using these algorithms can be targeted for identification in subsequent LC-MS/MS experiments using inclusion lists. These algorithms facilitated the identification of biologically relevant phosphopeptide isomers that could be missed in conventional phosphoproteomics analysis. This approach is complementary to data-dependant acquisition (DDA) and expands the scope of large-scale phosphoproteomics studies.

4.4. Materials and methods

4.4.1. Materials

Modified porcine sequencing grade trypsin was obtained from Promega (Madison, WI, USA). Acetonitrile (ACN) and HPLC grade water were purchased from Fisher Scientific (Whitby, ON, Canada). Ammonium bicarbonate and formic acid were obtained from EM Science (Mississauga, ON, Canada). Ammonium hydroxide, trifluoroacetic acid (TFA), DL-dithiothreitol (DTT), iodoacetamide, sucrose, MgCl₂, CaCl₂, Hepes, NaCl, glycerol, protease inhibitor cocktail (4-(2-aminoethyl)benzenesulfonyl fluoride [AEBSF], pepstatinA, E-64, bestatin, leupeptin, and aprotinin), phosphatase inhibitors cocktail (sodium vanadate, sodium molybdate, sodium tartrate, and imidazole), serum-free and protein-free insect medium were purchased from Sigma-Aldrich (Oakville, ON, Canada). Bradford protein assay was obtained from Bio-Rad (Mississauga, ON, Canada). Lactic acid was purchased from Fluka (Saint Louis, MO). Titanium dioxide bulk (5 µm, 500 mg) was obtained from Canadian Life Science (Peterborough, ON). Bond breaker tris[2-carboxyethyl] phosphine (TCEP) was purchased from Pierce Biotechnology Inc. (Rockford, IL). Phosphate buffered saline (PBS) and ethylene-diamine-tetra-acetic acid (EDTA) were obtained from HyClone (Thermo Scientific, Logan, UT). Oasis HLB cartridges (1 cc, 30 mg) were purchased from Waters (Milford, MA). Capillary columns for LC-MS were packed in-house using Jupiter C₁₈ (3 µm, 300 Å) particles from Phenomenex (Torrance, CA), and fused silica tubing from Polymicro Technologies (Phoenix, AZ). Synthetic phosphopeptides were purchased as Spike Tides from JPT Peptides technologies (Berlin, Germany).

4.4.2. Cell culture and protein extraction

Drosophila melanogaster Schneider S2 cells were cultured in serum-free and protein-free insect medium at 28°C. Cells were plated in 100 mm Petri dishes at a density of 18 millions of cells per dish. J774 (murine macrophage cells) were cultured in Dulbecco's modified Eagle's medium, pH 7.4, supplemented with 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine, and antibiotics (100 mg/mL penicillin and 100 mg/mL streptomycin) at 37 °C in a 5% CO₂ atmosphere. Cells were plated in 100 mm Petri dishes at a density of 10 millions of cells per dish. Intestine epithelial rat cells (IEC-6) were grown to confluence in 150 mm Petri dishes and then made quiescent by serum starvation for 24 h as described previously[199]. Total cell lysates were obtained using a detergent-free cell lysis. Cells were washed with PBS and cell lysis was performed in an extraction buffer (20 mM Hepes, 1.5 mM MgCl₂, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol) after sonication to extract cytoplasmic and nuclear proteins. After centrifugation at 13 000 g, the supernatant was isolated and the membrane pellet was discarded. During buffer preparation, 1mM DTT, proteases and phosphatase inhibitors cocktail were added. Total protein amount was quantified using Bradford protein assay.

4.4.3. Trypsin digestion

Proteins were reduced in 50 mM ammonium bicarbonate containing 0.5 mM TCEP and 1% SDS for 20 min at 37°C, and then alkylated in 5 mM iodoacetamide for 20 min at 37°C. Excess iodoacetamide was neutralized using 5 mM DTT. After dilution to 0.1% SDS with 50 mM ammonium bicarbonate, proteins were digested overnight with sequencing grade modified trypsin (enzyme:protein ratio 1:50) at 37°C with high agitation. The digest was acidified with TFA, and then evaporated to dryness in a SpeedVac.

4.4.4. TiO₂ phosphopeptides enrichment

Phosphopeptides from protein extracts were enriched on TiO₂ micro-columns as described previously[1]. Briefly, tryptic peptides were resuspended in 250 mM lactic acid (3% TFA/70% ACN) before loading on micro-columns pre-equilibrated with 3% TFA/70% ACN. A total of 750 µg of protein digest was loaded on one micro-column containing 6 mg of TiO₂ affinity media. Each micro-column was washed with the lactic acid solution followed by 3% TFA/70% ACN to prevent non-specific binding. Phosphopeptides were eluted with 1% NH₄OH pH 10 in water and then acidified with TFA before desalting on Oasis HLB cartridges. Samples were evaporated to dryness on a SpeedVac and the equivalent of 250 or 1500 µg of starting material was resuspended in 20 µL of 0.2% formic acid/5% ACN for 1D or 2D-LC-MS/MS experiments, respectively.

4.4.5. Mass spectrometry

All LC-MS/MS experiments were performed on an Eksigent 2D nano-LC (Dublin, CA) system coupled to a LTQ-Orbitrap XL or Velos mass spectrometers (Thermo Scientific, Bremen, Germany). Custom reversed-phase trap (4 mm length, 360 µm i.d.) and analytical columns (15 cm length, 150 µm id) were packed with Jupiter C₁₈ stationary phase (3 µm particle, 300 Å pore size). Chromatographic separations of enriched phosphopeptide extracts were achieved using a flow rate of 600 nL/min and a linear gradient from 5-30% ACN over of 70 or 120 min. Online 2D-LC separations were performed using an Opti-Guard SCX column (1 mm length, 350 µm id, Optimize Technologies) and eluted with 0, 0.25, 0.5, 0.75, 1.0, and 2.0 M ammonium acetate salt fractions, pH 3 (5% ACN, 0.2% FA). The survey scans were acquired at a resolution of 60 000 (AGC: 10⁶, injection time: 500 ms) over the acquisition range of *m/z*: 300-2000. MS/MS spectra with multistage activation were acquired in CID mode using data-dependent acquisition for multiply charged ions exceeding a threshold of 10 000 counts (isolation width: 2 Da, normalized collision energy: 35, activation q: 0.25, activation time: 30 ms, dynamic exclusion: 30 s). MS/MS spectra acquired using ETD mode were obtained using 100 ms activation time with supplemental

activation. HCD MS/MS spectra were obtained using 0.1 ms activation time. In each cycle, 12 MS/MS scans on the most abundant peptides were performed in the LTQ (AGC: 5×10^4 , maximum injection time: 300 ms) for CID/ETD and 6 MS/MS for HCD. Targeted analyses were achieved using an inclusion list generated by custom algorithms detailed in the **section 4.4.7**. MS/MS were triggered on selected parent masses with ± 0.02 Da tolerance. To confirm the identification of MS/MS spectra of phosphopeptides isomers, repeated precursor selection was performed every 30 or 5 seconds for data-dependant acquisition (DDA) and targeted analyses, respectively.

4.4.6. Database search

MS data were analyzed using the Xcalibur software (version 2.0.7). Mascot Distiller 2.3.2 was used to generate MS/MS peak-list files from raw data. For the detection of phosphoisomers, MS/MS spectra from the same precursor m/z were not combined together. All MS/MS spectra were searched with Mascot 2.3 (Matrix Science) against a concatenated target/decoy database of the relevant specie (fly: Uniprot v11.0 - 29 447 proteins, mouse: IPI v3.54 - 55 987 proteins, rat: IPI v3.54 - 39 928 proteins) using the following parameters: peptide mass tolerance ± 10 ppm, fragment mass tolerance ± 0.5 Da for CID/ETD and 0.02 Da for HCD, trypsin enzyme with up to 4 missed cleavages, and carbamidomethyl (C), deamidation (NQ), oxidation (M), phosphorylation (STY) as variable modifications. Finally, ProteoConnections was used to constrain the peptide false discovery rate (FDR) to less than 1% and to assign the phosphorylation site localization confidence [199].

4.4.7. Algorithms for detecting phosphopeptide isomers from LC-MS/MS analyses

Two algorithms were developed to detect potential phosphoisomers from LC-MS elution profiles and from mixed MS/MS spectra. The algorithms were implemented in C# and run under Windows operating system. The detection of phosphopeptide isomers from LC-MS elution profile comprises four steps (**Figure 4.2. A**): *i*) generation of peptides map, *ii*) selection of identified phosphopeptides, *iii*) search for isobaric peptides and *iv*) output list of candidates to file. Step *i* and *ii* are pre-requisite for step *iii*. Step *i* detects peptide features from raw LC-MS data files. Peptide detection is performed using an in-house version of the THRASH algorithm 28. It creates a peptides map with the following peptide features: m/z , retention time (minimum, peak top, maximum) and peak intensities for all elution tracks of each peptide isotope. Step *ii* consists of selecting phosphopeptides that can have more than one possible phosphorylatable residue (serine, threonine and tyrosine amino acids). Identified phosphopeptides attributes (m/z , retention time, peptide sequence and modifications) are loaded from ProteoConnections output CSV file. Step *iii* uses both input files to find isobaric peptides. For each identified phosphopeptide, the algorithm searches the peptides map for neighboring phosphopeptide isomers within a retention time window $rtWin$ of 4 minutes and an m/z tolerance $mztol$ of 10 ppm with identical charge state z . Next, the algorithm compares the isotopic profiles of candidates to detect phosphopeptide isomers separated by a minimum valley $minRtValley$ of 0.1 min. Multiple replicate LC-MS analyses can be used to report reproducibly detected phosphopeptide isomer candidates and reduce false positive identifications arising from inconsistent peak shape. Finally in step *iv*, candidates are reported in an inclusion list (target m/z , minimum and maximum retention times). The inclusion list is generated in a text file format compatible with the acquisition method interface of the Thermo Xcalibur software. The inclusion list can be separated into multiple files with appropriate number of precursor ions if the number of candidates exceeds the physical limit of data acquisition of the instrument. A text report is also generated for review and editing before the launch of the targeted analysis.

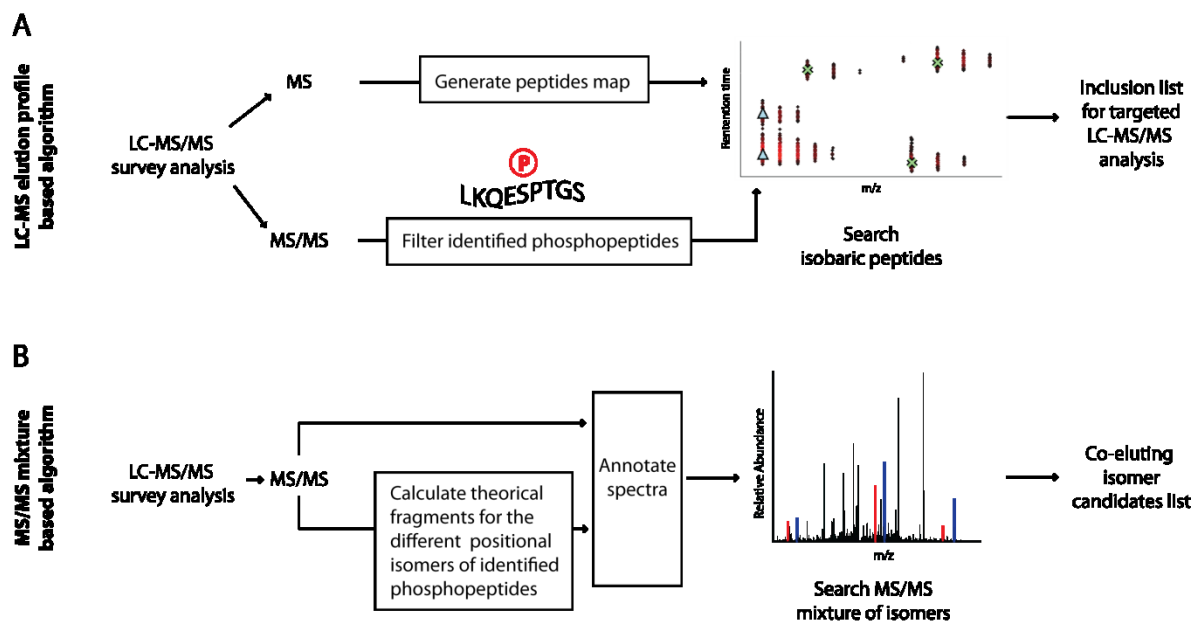


Figure 4.2. Algorithm workflow for detecting phosphopeptide isomers.

Two algorithms were implemented to search for phosphopeptide isomers. A. Search for partially or fully separated phosphopeptide isomers from LC-MS elution profile. B. Detection of phosphopeptide isomers from mixed MS/MS spectra.

The algorithm for the detection of phosphopeptide isomers from mixed MS/MS spectra searches for site specific fragment ions (**Figure 4.2. B**), and comprises 4 steps: *i*) calculation of phosphopeptide fragments for each positional isomer, *ii*) fragment ions annotation, *iii*) search of site-specific fragment ions for each positional isomer, and *iv*) report of potential phosphopeptide isomer candidates. In step *i*, all phosphorylatable residues are identified to generate all possible phosphorylated combinations. Theoretical fragment ion patterns are generated for all combinations using Peptide Fragmentation Modeller software 29 to create a list of site-specific fragment ions. In step *ii*, each peak p_i from a spectrum s is annotated using the theoretical fragments of each positional isomer based on user's specified fragment mass tolerance Δm . The presence of multiple isomers is detected in step *iii* by searching pairs of fragment ions suggesting the occurrence of more than one positional isomer. These represent pair of unique fragments to each isomer,

partially shared fragments (indicate that at least two forms are present but cannot be distinguished), or a shift of 80 Da for a specific fragment (see **Table 4-I**). Finally, a list of co-eluting isomer candidates is generated in step *iv*. The application also contains an automated validation software tool to review the isobaric phosphopeptide candidates of a targeted LC-MS/MS analysis. For each candidate, a text report indicates all acquired MS/MS spectra and the associated peptide identification when successfully assigned by Mascot. Candidates sharing the same sequence but differing in the site of the phosphorylated residues are highlighted as phosphopeptide isomers using a simple binary classifier based on retention time. The selection of candidates for targeted LC-MS/MS analysis involves the following steps: *i*) peptide identification with good score and confident site localization are retrieved for each acquired MS/MS spectrum, *ii*) peptide sequences are compared and non-identical ones are discarded, *iii*) peptides phosphorylated at different positions are retained, and *iv*) a binary classification of MS/MS identifications is performed based on retention time separating isomer candidates. The binary classification finds the optimal retention time threshold that separate MS/MS of isomers based on their elution order. The accuracy of classification is then calculated and candidates above a user defined threshold are reported as positional isomers.

4.4.8. Datasets availability

All datasets generated or used in this study are available online via ProteoConnections at <http://www.thibault.irc.ca/proteoconnections>. For the rat phosphoproteomics dataset, a description of the experimental and analytical procedures are available in Courcelles *et al* [199]. All phosphorylation sites are available in **Table A4- I**.

Table 4-I. Types of fragment ions that reveal the presence of phosphopeptide isomers.

Fragment m/z	Fragment type	RKp S EEPTAPSGNK	RKSEEP p TAPSGNK	RKSEEP TAP pSGNK	
302.36	–	c2	c2	c2	
388.32	Partially shared	z4	z4		
389.23	Shifted		c3		
469.28	Shifted	c3			←
598.33	Unique	c4			←
657.44	Unique	z7			
824.43	Unique	c6			
925.48	Partially shared	c7	c7		
963.41	Partially shared		z9	z9	
1013.36	Shifted			c9	←
1092.41	Partially shared		z10	z10	←
1093.53	Shifted	c9	c9		←
1179.43	–	z11	z11	z11	
1180.45	–	c10	c10	c10	

Unique: Peak assigned to a fragment specific to one phosphopositional isomer

Shifted: Pair of peaks separated by 80 Da assigned to the same fragment that indicates the presence of two phosphopositional isomers.

Partially shared: Peak assigned to a fragment that is shared by few phosphopositional isomers. Compared with unique or other partially shared fragments and if the is no overlap, it can reveal if more than one isoform is truly present.

– : Not informative to distinguish isomers.

4.5. Results and Discussion

4.5.1. Occurrence of phosphopeptide isomers in large-scale phosphoproteomics studies

The presence of phosphopeptide isomers has been described only in few reports for specific phosphoproteins, and limited information is currently available on their distribution in large-scale phosphoproteomics studies[378-383]. Their detection by MS can be problematic depending of the MS/MS acquisition parameters used. For example, dynamic exclusion for period extending to 1-2 min could prevent the identification of closely eluting phosphopeptide isomers. Also, the grouping of MS/MS spectra often applied to simplify data processing would prevent their successful differentiation by search engines.

To gain further insights on the occurrence of phosphopeptide isomers from large-scale proteomics datasets, we developed two algorithms that detect separated and co-eluting phosphopeptide isomers (**Figure 4.2**). The first algorithm utilizes peptide maps and MS/MS spectra to correlate assignments and proximity of candidates (**Figure 4.2. A**). The tolerance on mass and retention time can be defined by the user, and typically retain assigned phosphopeptides that are within 10 ppm of the calculated mass and separated by at least 0.1 min. Only peptides with a Mascot score higher than 20 and site localization confidence greater than 75% are retained.

We used this detection algorithm to determine the overall distribution of phosphopeptides from three different large-scale phosphoproteomics studies of J774 mouse macrophages, *Drosophila melanogaster* S2 cells and a previously reported dataset from a rat intestinal epithelial cell line IEC-6[199]. In all cases, phosphopeptides were enriched using TiO₂ media and analyzed by online 2D-LC-MS/MS on an LTQ-Orbitrap XL. **Table 4-II** reports the distribution of identification for all three datasets, where 4 798, 15 700, and 10 110 unique phosphopeptides were identified in J774, IEC-6 and S2 cells, respectively. For each dataset, we next evaluated the number of potential phosphopeptide isomers that are represented by at least two separate LC peaks. These analyses with our algorithm revealed that approximately 8-12% of phosphopeptides have potential isomeric counterparts based on their LC-MS elution profile. However, no more than 20% of these candidates were potential isomers based on the best match found by MS/MS. Closer examination of these MS/MS spectra indicated that only a subset of these identifications corresponded to true phosphopeptide isomers due to insufficient time separation and/or ambiguous site localization. For example, we detected 1 298 phosphopeptide isomer candidates with the algorithm in the IEC-6 dataset, but only 48 out of the 145 isomers suggested by the search engine were true isomers with high localization confidence. The analysis of these large-scale phosphoproteome datasets indicated that 3-6% of all identified phosphopeptides were confirmed as true isomers based on high confidence site assignment.

Table 4-II. Relative distribution of phosphopeptide isomers from large-scale phosphoproteomics studies.

Cell extract	# unique peptides	# unique phosphopeptides	# phosphate per peptide			# sites with high confidence	Phosphorylated residues			# detected phosphopeptide isomers	
			1	2	≥ 3		Ser	Thr	Tyr	LC	LC + MS/MS
J774	5372	4798	4071	669	58	3214	3598	960	240	598	53 (19)*
IEC-6	23902	15700	12717	2669	314	10676	12560	2826	314	1298	145 (48)*
S2	13699	10110	7926	1739	445	7583	7785	1921	404	1304	276 (117)*

All datasets are available online via ProteoConnections at <http://www.thibault.irc.ca/proteoconnections>. Data for the rat IEC-6 phosphoproteomics analyses were described in Courcelles, *et al.*[199]. Numbers indicated by * are confirmed isomers with a high confidence in phosphosite localization.

However, this number likely underestimates the real distribution of phosphopeptide isomers due to the stochastic sampling of precursor ions based on their abundance and the potential co-elution of isomeric candidates.

4.5.2. Correlation of physical properties of phosphopeptides isomers

Chromatographically separated phosphopeptide isomers from mouse and rat were combined into a single list of 64 isomeric congeners to study their intrinsic physicochemical properties (**Table A4- II**). The length of separated phosphoisomers ranged from 7 to 35 amino acids (average of 17), and the distance between phosphorylated sites spanned from 1 to 23 (average of 15) (**Figure 4.3. A-B**). These results are consistent with isomers reported in the literature where their length varied from 13 to 26 amino acids, while phosphorylated sites were separated by 1 to 16 amino acids[378-383]. The measured shifts in retention time caused by the alternate modification site varied from 0.5 min to 6 min (outlier at 22.3 min) with an average of 2.3 min (**Figure 4.3. C**) and LC peak resolution ranged from 1 to 12.8 with an average of 4.6 (**Figure 4.3. D**). In an attempt to correlate changes in retention times with the sites of phosphorylation, we considered prediction algorithms such as BioLCCC[390]. Unfortunately, we obtained mixed success in accurately predicting retention times of monophosphorylated peptide isomers using this algorithm. This could be partly explained by potential changes in 3D peptide conformations that are difficult to simulate when retention time modeling is based only on phosphorylated peptides and their non-phosphorylated counterparts.

To rationalize the difference in retention times between isomeric phosphopeptides, we calculated the corresponding changes in local hydrophobicity and electrostatic interactions. The difference in local hydrophobicity scale was calculated using a windows of 3 amino acids centered on the phosphosite and the amino acid scale values from Kyte and Doolittle[391] (**Figure A4. 1. A**). Electrostatic interactions were determined by calculating the difference in distance between the site of phosphorylation and the charged residues (**Figure A4. 1. B-C**). We also evaluated the potential electrostatic interactions of the phosphorylated residues with the free α -amino group of the N-terminal amino acid.

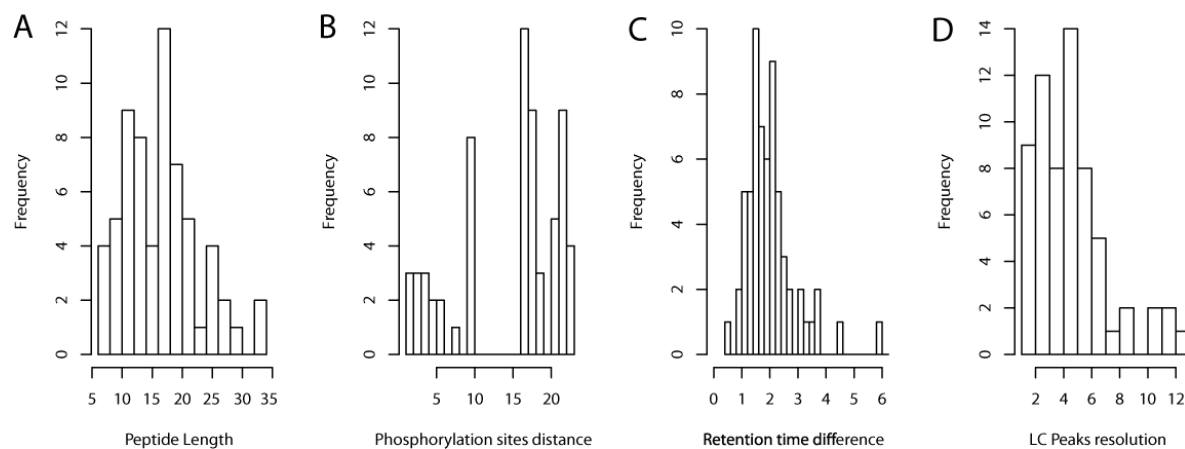


Figure 4.3. Physico-chemical properties of phosphopeptide isomers separated by reverse phase LC.

Properties of 64 phosphopeptide isomers identified in IEC-6 and J774 cells using the LC-MS elution profile detection algorithm and MS/MS identification. Histograms representing (A) the distribution of the peptide length of phosphopeptide isomers, (B) distance in amino acids between phosphorylated residues, (C) difference in retention time between phosphopeptide isomers (an outlier with 22.3 min of separation is not displayed for clarity), and (D) chromatographic resolution of isomeric pairs.

While no linear relationship could be established between the number of amino acid separating the phosphorylated residue from the N-terminus and retention time, we noted that 73% of phosphopeptides with modified residues further away from the N-terminus eluted first (**Figure A4. 2**). In particular, pairs characterized by phosphorylated residues proximal and distant from the N-terminus exhibited the largest shift in retention times. This observation is reminiscent of the change in retention time observed between phospho- and dephosphopeptide counterparts where increased hydrophobicity of phosphopeptides was attributed to ion-pairing interactions with the neighboring N-terminus amino group[243]. An example is shown in **Figure 4.4** for the separation of two phosphopeptides from δ -catenin. In this case, the alternate phosphorylation of adjacent Ser residues from tryptic peptide VGGSSVDLHR resulted in two peaks separated by 1.2 min (**Figure 4.4. A**). The localization of both phosphorylation site was determined with high confidence (**Figure 4.4. B-C**), and unambiguously identified them as distinct phosphopeptide isomers.

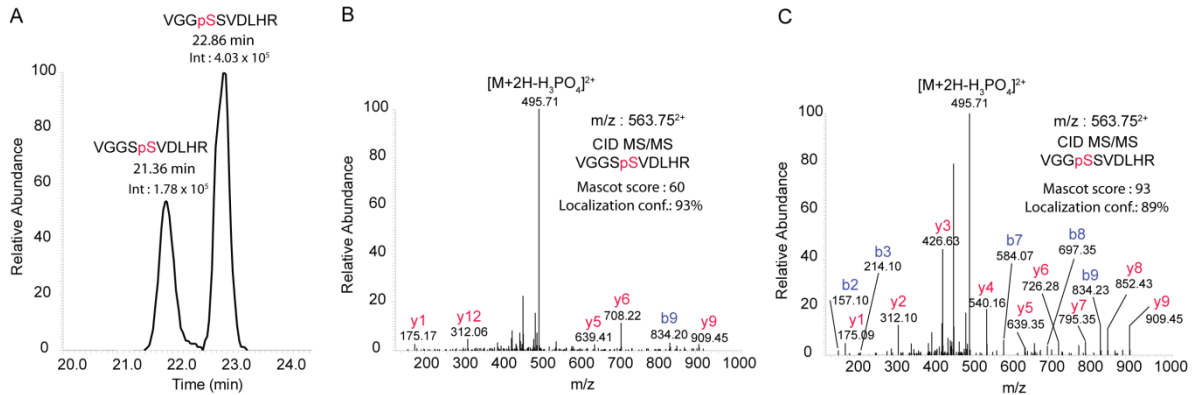


Figure 4.4. LC-MS separation of two isomeric phosphopeptides.

A. Extracted ion chromatogram for the doubly-protonated phosphopeptides VGGSpSVDLHR and VGGpSSVDLHR at m/z : 563.75²⁺, from a TiO₂-enriched digest of IEC-6 cell extract showing a 1.2 min separation between the two phosphopeptide isomers. MS/MS spectra of peaks eluting at 21.4 (B) and 22.9 min (C).

The phosphorylated Ser residue proximal to the N-terminus was more hydrophobic and eluted last consistent with that observed previously (**Figure A4. 2**). Additional studies using peptide conformation modeling software are thus warranted to investigate the 3D structural changes imparted by the localization of phosphorylate residue and their impact on retention time.

4.5.3. Analysis of synthetic phosphopeptide isomers

We evaluated the ability of our algorithm to detect chromatographically separated isomers using a set of nine synthetic peptides phosphorylated at different positions. These phosphopeptides were selected from candidates previously identified in a large-scale phosphoproteomics study of *D. melanogaster* S2 cells[1]. We obtained monophosphorylated peptides for all phosphorylatable residues of each peptide for a total of 26 different phosphopeptides. Three peptides were phosphorylated at two positions, four at three, and two at four positions. Phosphopeptides were analyzed by LC-MS/MS either individually or mixed together in different combinations. We performed LC-MS/MS

experiments on the LTQ-Orbitrap XL using an inclusion list to trigger the acquisition of CID and ETD MS/MS spectra of target synthetic phosphopeptides every 15 seconds.

When analyzed individually, all 26 monophosphorylated peptides were identified and localized with high confidence (> 0.75) using either ETD or CID, confirming the expected phosphorylation sites (**Table A4- III**). LC-MS/MS analyses of phosphopeptides mixed in different combinations indicated that isomers were separated with variable degree of overlap, and only 18 were clearly separated from each other. An example is shown in **Figure 4.5** for the separation of SLGANSSTDNVDGKK phosphorylated at either Ser1, Ser6, Ser7 or Thr8 residues. In this case, two of the four phosphopeptides were completely resolved while monophosphorylated peptides at Ser1 and Ser6 co-eluted together (**Figure 4.5. A**). MS/MS spectra obtained from phosphopeptides analyzed individually provided confident assignment for each isomer (**Figure 4.5. B-E**).

We next evaluated the performance of our algorithm to identify potential phosphopeptide isomers. Our algorithm could detect meaningful separation for most phosphopeptide isomers, except for 8 co-eluting phosphopeptide isomers. The algorithm also found more phosphopeptide peaks than expected, and close examination of the corresponding MS/MS spectra revealed almost superimposable fragmentation patterns, suggesting the presence of separated phosphopeptide conformers as previously reported for synthetic peptides[392]. These conformers were wrongly assigned as potential isomer candidates and were undistinguishable from their LC-MS elution profiles alone. The presence of phosphopeptide conformers could lead to ambiguous isomer assignment if they occur frequently in protein digests.

We then applied our validation algorithm to confirm phosphopeptide isomers from their corresponding CID and ETD MS/MS spectra. Unambiguous identification were obtained for 18 phosphopeptide isomers, while the algorithm could not determine the phosphorylation sites with high confidence for 8 isomeric congeners irrespective of the activation mode used. The low abundance of site-specific fragment ions can prevent unambiguous identification of phosphorylation site, and represent a significant obstacle to distinguishing phosphopeptide isomers[280].

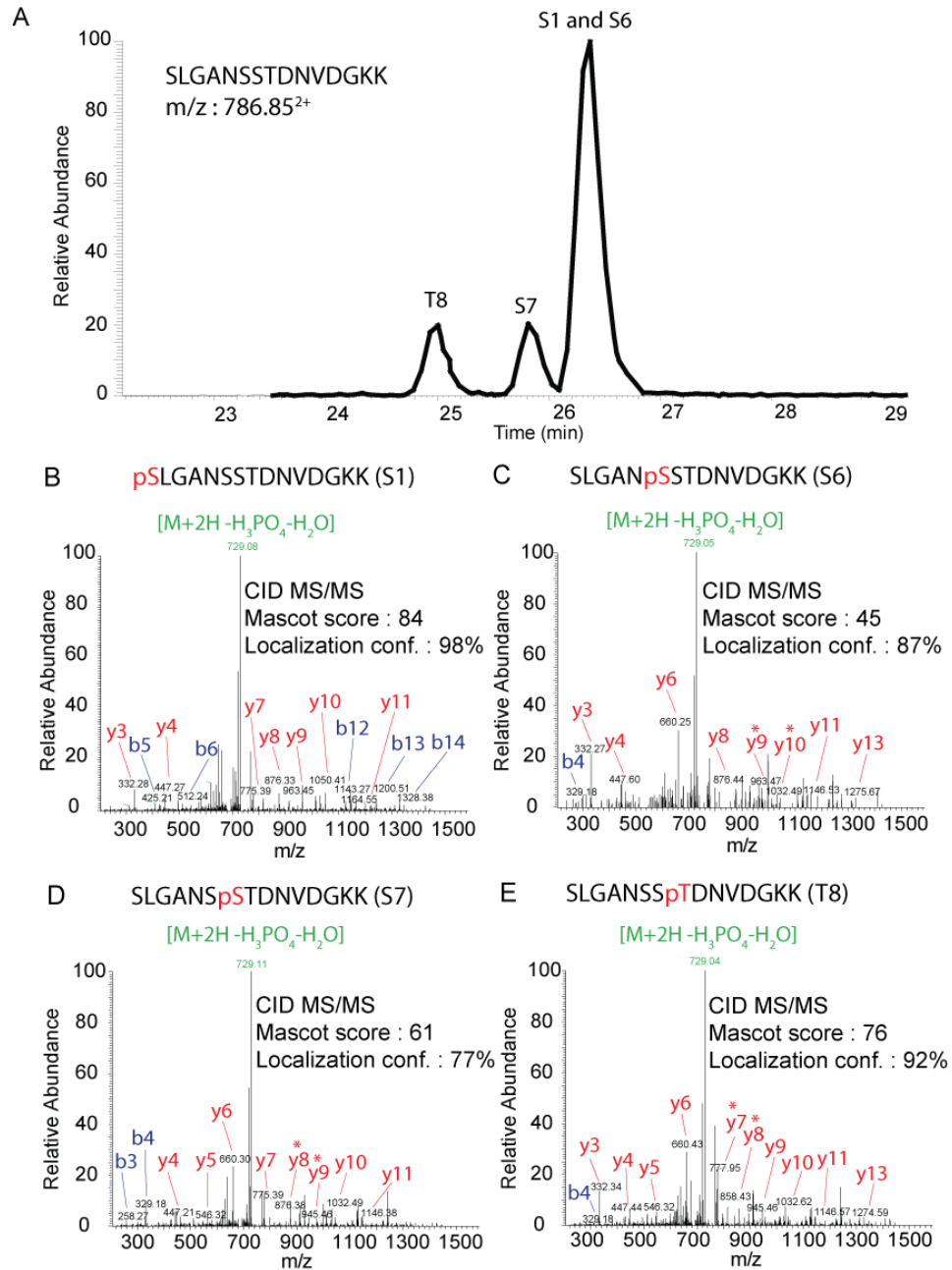


Figure 4.5. Detection of synthetic phosphopeptide isomers for the peptide SLGANSSTDNVDGKK using the algorithm based on LC-MS elution profile.

A. LC-MS profile of the extracted ion m/z : 786.85²⁺ showing separated isomers (Ser7 and Thr8) and 2 co-eluting isomers (Ser1 and Ser6). CID MS/MS spectra confirming the phosphorylated Ser1 (B), Ser6 (C), Ser7 (D) and Thr8 (E).

Partially separated phosphopeptide isomers can provide distinct MS/MS spectra if the acquisition was triggered in a region with limited overlap. These experiments indicated that the algorithm can detect isomers showing partial separation, although some assignments could be misrepresented if phosphorylated sites are ambiguously localized.

4.5.4. Targeted analysis of phosphopeptide isomers

Our original 2D-LC-MS/MS on the LTQ-Orbitrap Velos using CID, ETD and HCD activation modes enabled the identification of 10 110 unique phosphopeptides from *D. melanogaster* S2 cells, of which 4 783 had sites localized with high confidence (see **Figure A4. 3. A** for the contribution of each activation modes). Comparison of acquired MS/MS spectra confirmed 77 pairs of phosphopeptide isomers and 199 ambiguous pairs with uncertain localization (**Table 4-II**). The algorithm detected 1 304 phosphopeptide isomer candidates that showed 2 peaks separated by at least 0.1 min (**Table A4- IV**). We next performed a 2D-LC-MS/MS analysis using a target list of phosphopeptide candidates. This list was split in two separate runs to maximize the capture of targeted phosphopeptide isomers. Each precursor ion was fragmented sequentially by CID, ETD and HCD every 5 seconds to obtain good quality MS/MS spectra, and to confirm the localization of phosphorylation site with high confidence. We identified a total of 1 033 unique phosphopeptides from these experiments, of which one third were identified in all three fragmentation modes (**Figure A4. 3. B**). Out of these identifications, we obtained a total of 86 true isomers with high localization confidence, and 216 phosphopeptide pairs with low confidence (< 0.75). A subset of 46 true isomers was also identified by DDA while 40 were unique to the targeted analysis. The 31 isomers identified only by DDA were all detected by the algorithm but were missed in the targeted analysis due to ambiguity in their assignment or insufficient duty cycle to capture all isomeric phosphopeptides for each fragmentation mode. Altogether, the DDA and targeted analyses identified a total of 117 true phosphopeptide isomers corresponding to 1.2 % of all phosphopeptides (**Table A4- IV**). The targeted analysis yielded complementary results to the DDA analysis with

approximately 50% overlap of identification (**Figure A4. 4**). A comparison of physicochemical properties of phosphopeptide isomer from S2 cells (**Figure A4. 5**) with those identified in J774 and IEC-6 cells (**Figure 4.3**) revealed similar trends, except for the occurrence of relatively long peptides (> 35 a.a.) observed in S2 extracts.

We also observed non-phosphorylated peptides with m/z and retention time similar to those of phosphopeptide isomers that could be misidentified as potential isomeric congeners (**Figure A4. 6**). Those artifacts represented 7% of all included peptide ions. Among these, we noted 8 cases with alternative tryptic cleavage sites, 56 with alternative peptide sequences, 2 with a change in the position of an alternate modification, and 20 corresponding to phosphopeptide conformers. Alternative tryptic cleavages occurred when two arginine or lysine residues were found at both C- and N-terminus (e.g. XR↓KPEPTIDER↓KX ↔ XRK↓PEPTIDERK↓X, where trypsin cleavages occur at ↓). While this artifact can be anticipated from the protein sequence, alternate missed cleavages were observed only for 15% of all predicted cases. Phosphopeptide conformers were present at a low frequency in protein extracts. An example of this is shown in **Figure A4. 7** for the separation of two conformers for the phosphopeptide GIMEEIEMRpSPLSDR (m/z : 625.6³⁺). Both conformers are well resolved by reverse-phase LC and have a common HCD fragmentation pattern that localizes the phosphorylated site on the same residue. The second eluting form has an additional fragment ($y9^{++}$) that could be favored by an alternative conformation. These conformers could possibly arise from cis-trans proline isomerization or from deamidation of asparagine or glutamine residues into their corresponding isomeric acids[393]. To determine the occurrence of isomeric peptides, we applied our detection algorithm to LC-MS data from tryptic digest of a mouse macrophage J774 protein extract. We detected 127 potential isomeric peptide pairs from a total of 1 869 identified peptides. Almost half of these potential isomers (73 pairs) showed alternate peptide sequences (data not shown) while the remainder (54 pairs) were identified with the same sequence. However, we could obtain meaningful identification for only 3 of the 54 pairs. These observations indicate that changes in the peptide backbone (e.g. deamidation, isomerization, etc...) can also result in shift of retention time that further complicate the

interpretation of phosphopeptide isomers. In addition, we identified 7 phosphopeptides with three isomers and one phosphopeptide with four isomers (**Figure 4.6**). The sequence of the latter peptide IPSSSSDFSK from a protein of unknown function contains 5 phosphorylatable residues, of which 4 were identified in our analysis. Interestingly, the localization of the phosphorylation site had a significant influence on the phosphopeptide hydrophobicity as shown for phosphorylated Ser3 and Ser4 residues that are separated by 3.3 min (**Figure 4.6. A**). Phosphopeptides bearing modification at Ser5 and Ser6 residues were unresolved and co-eluted with the former two phosphopeptide isomers. We next compared the confidence in the site localization for all MS/MS spectra acquired over this time period (**Figure 4.6. B** and **Figure A4. 8**). While phosphorylation at Ser3 and Ser4 residues was localized confidently using all three fragmentation modes, the localization of Ser5 was confirmed only by HCD/CID and Ser6 by ETD. This result indicates that multiple fragmentation methods are required to identify all positional isomers of a peptide. This peptide can also be phosphorylated on Ser9 but no MS/MS spectrum supported this possibility. Repeated MS/MS acquisition were required to confirm the presence of each phosphorylated isomer in order to obtain a consistent fragmentation pattern for the identification of site-specific modifications (**Figure 4.6. C**). The use of repeated MS/MS acquisition enabled the profiling of fragment ion patterns during the elution to increase the localization confidence of phosphorylation sites of isomeric phosphopeptides.

4.5.5. Detection of co-eluting phosphopeptide isomers

The analysis of chromatographically separated phosphopeptide isomers revealed that the difference in retention times between isomeric pairs extends from 0.5-7.0 min, and more than 56% were observed within 2 min of each other. The close proximity of retention times between isomeric pairs raised the possibility that a sizable proportion of phosphoisomers could co-elute together. To evaluate this possibility, we extended the functionality of the present algorithm to detect isomers from fragment ion features of the MS/MS spectra.

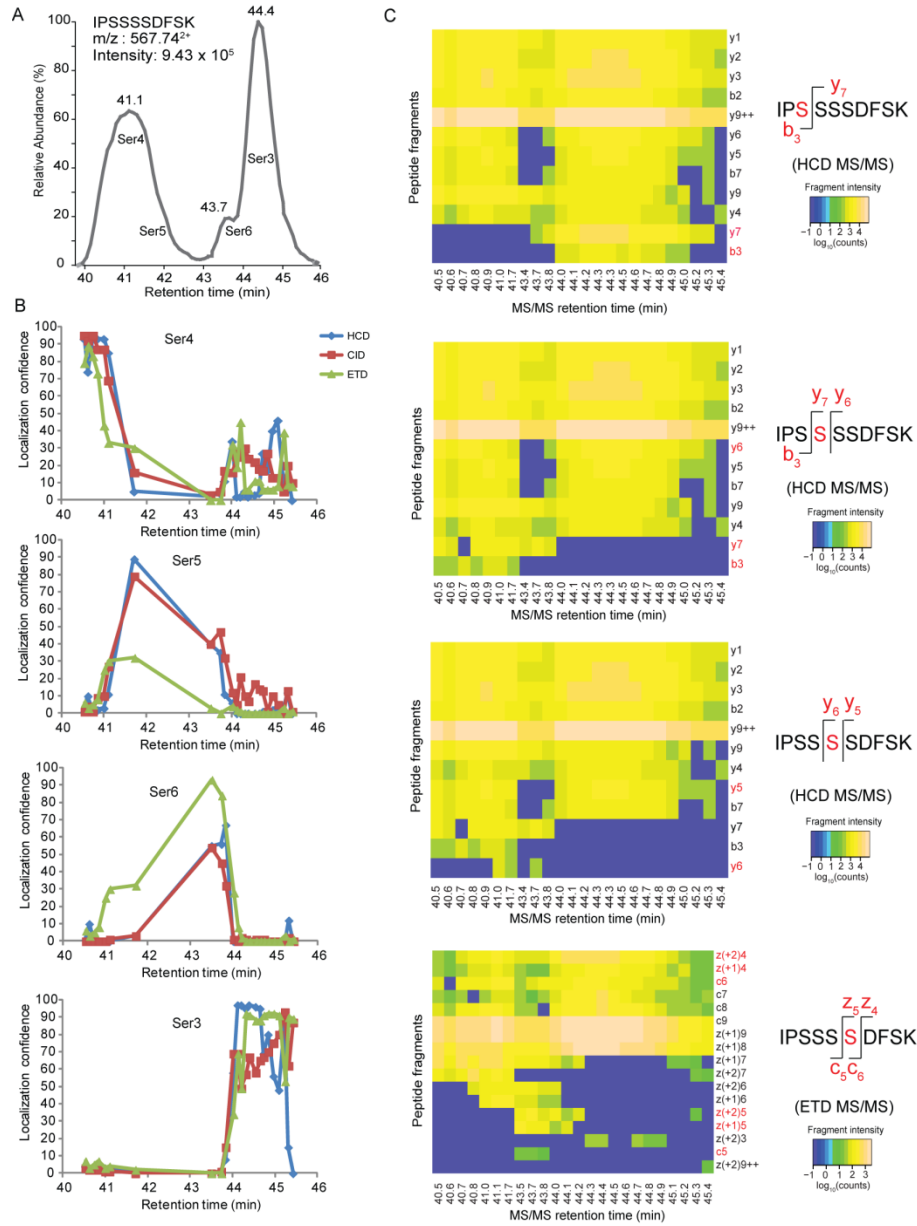


Figure 4.6. Identification of four phosphopeptide isomers of IPSSSSDFSK by targeted LC-MS/MS analysis.

A. Extracted ion chromatogram for m/z : 567.742⁺ for monophosphorylated peptides from IPSSSSDFSK of TiO₂-enriched phosphopeptides from *D. melanogaster* S2 cells. B. Localization confidence (%) vs. retention time for all MS/MS acquired in HCD, CID and ETD fragmentation modes. C. Heatmap representing each phosphorylated position according to the intensity of fragment ions. Fragments enabling site-specific localization are highlighted in red.

Enriched phosphopeptides from *Drosophila melanogaster* S2 cell line were analyzed by 1D-LC-MS/MS on the LTQ-Orbitrap Velos using CID, ETD and HCD activation modes. A list of potential phosphopeptide isomers was generated from a preliminary 1D-LC-MS/MS analysis obtained using DDA. A dynamic exclusion of 10 s was used to acquire multiple MS/MS spectra across the peak elution and identify co-eluting isomers partially separated by LC. A total of 2 901 unique phosphopeptides were identified, of which 1 544, 1 251 and 861 phosphopeptides were assigned from their corresponding CID, ETD and HCD MS/MS spectra.

The algorithm annotated each MS/MS spectrum based on theoretical fragment ions calculated for each possible isomer. A search was then initiated to identify fragment ion pairs that distinguish two or more phosphopeptide isomers. The distribution of different types of fragment ions observed between phosphopeptide isomers is presented in **Figure 4.7**, and indicated that the number of observed fragment ions enabling isomer detection depends on the fragmentation method used. For an arbitrary cut-off of at least two fragments pair, the proportion of candidate isomers co-eluting was 65% (1 011/1 544), 9.8% (123/1 251), 3.6% (31/861) for CID, ETD and HCD, respectively. MS/MS spectra acquired with HCD generate a lower proportion of isomer candidates than other fragmentation modes. The false positive rate is lower due to enhanced mass accuracy for fragment ion detection. An additional problem with CID fragmentation in the LTQ is the potential gas phase rearrangement of the labile phosphate group[394]. This phenomenon generates artificial isomers that could be misidentified by the algorithm. Visual inspection of elution profile of ETD and HCD isomer candidates confirmed that 23 pairs of phosphopeptide isomers were partially resolved. This result suggests that co-eluting phosphopeptide isomers represent at least 0.8% of all identified phosphopeptides.

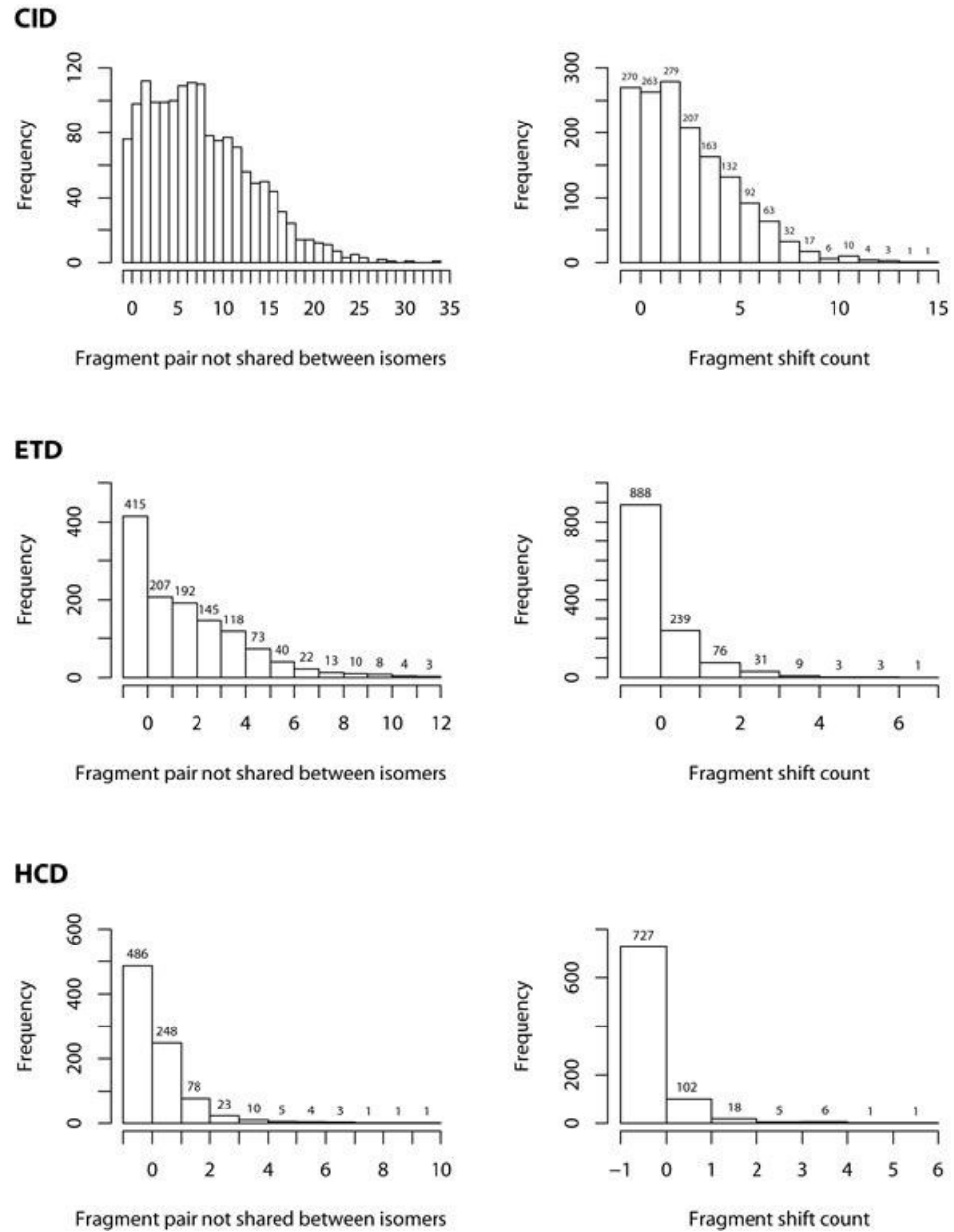


Figure 4.7. Detection of co-eluting phosphopeptide isomers using distinct fragment ion features for MS/MS spectra acquired with different activation modes.

The MS/MS based algorithm was tested using three different fragmentation methods (CID, ETD and HCD) to detect co-eluting phosphopeptide isomers. Histograms show specific characteristics used to detect isomers from MS/MS spectra acquired by different fragmentation modes.

4.5.6. Algorithm applications

The algorithms developed as part of the present study were primarily used for the detection of phosphopeptide isomers to determine the extent of their distribution in large-scale phosphoproteomics experiments, and to ascertain the localization of ambiguous phosphorylation sites. Their practical applications also extend to the identification of phosphorylation sites that could elude traditional data mining approaches used in phosphoproteomics. The relatively low occurrence of phosphopeptide isomers in TiO₂-enriched extracts (typically 3-6% of all phosphorylation sites) dictates the use of targeted strategies for their detection. Obviously, the detection of phosphopeptide isomers required additional targeted LC-MS/MS analyses with inclusion lists generated from a preliminary DDA experiment. The number of targeted analyses must be evaluated based on sample availability and requirements for comprehensive identification of phosphopeptide isomers. The precise identification of phosphorylation site when several phosphorylatable residues are present on the same peptide is of particular importance in functional studies involving site-directed mutagenesis experiments. Indeed, large-scale phosphoproteomics experiments provide a wealth of information but also ambiguities on the identification of phosphorylation sites that need to be resolved before the launch of resource-consuming functional studies. In addition, these algorithms can also facilitate the quantitation of closely eluting phosphopeptide isomers that can be mistakenly combined together by search engines or quantitation softwares. For example, partially resolved isomers can be wrongly quantified in label-free quantification if time alignment between runs is imperfect. In this case, the quantitation software could match the abundance of different phosphopeptide isomers and provide the wrong change in phosphorylation between experimental paradigms. This potential difficulty is highlighted by our algorithms for future validation and review by the user.

Another application of these algorithms is for the detection of modified residues that could act as phosphorylation switch to regulate protein-protein interactions. The addition of a negatively charged phosphate group on amino acids can modulate intramolecular and intermolecular interactions. The formation of hydrogen bonds and electrostatic interactions

via phosphorylated residues can change protein conformations and their biological functions. A notable example is the activation of the insulin receptor tyrosine kinase where the phosphorylation of the activation loop impart a conformational change that confers access of substrates to the active site[395]. Interestingly, we noted that our datasets contained several phosphorylation sites located within a relatively short peptide sequence, suggesting that the local density of modified residues could mediate intramolecular interactions. To evaluate the overall distribution of phosphorylation sites in phosphoproteins from *D. melanogaster*, we combined our identification to those of other large-scale phosphoproteome analyses such as PHOSIDA[396] and PhosphoPep[284]. We determined the number of phosphorylated sites for different sequence length ranging from 5 to 25 amino acids, and found 56 sequence clusters that comprised 3 or more phosphorylation sites on segments shorter than 25 amino acids. An unusually high number of modified residues was found on SR family splicing factor SF2 (Q9V3W7), where 11 phosphorylation sites were identified on the C-terminus. This protein shares 56% sequence identity with the human serine arginine rich protein splicing factor 1 (SRSF1), a protein that participate in the maturation of the spliceosome. SR proteins are phosphorylated by two families of kinases SRPK and CLK to alter their cytoplasmic nuclear localization and splicing functions[397]. We also observed several proteins where phosphorylation sites were located close to nuclear localization sequence (NLS) or to poly-lysine or poly-arginine domains. For example, H/ACA ribonucleoprotein (RNP) complex subunit 4 contains 8 phosphorylated residues (position 440-460) that precede a C-terminal poly-lysine domain. This protein binds to RNA and plays a central role in ribosomal RNA processing[398]. The phosphorylation of residues close to poly-lysine domains could modulate electrostatic interactions to affect RNA binding and/or association with accessory proteins of the RNP complex. Moreover, Ser₄₄₂ was phosphorylated upon insulin stimulation (unpublished data) and this site is modified by p70S6K kinase based on NetworKIN[399]. Interestingly, Thr₄₄₃, Ser₄₄₄ and Ser₄₄₅ are phosphorylated by CK2 while Ser₄₄₉ is phosphorylated by shaggy (homologue of GSK3). We also identified two neighboring residues Ser₂₁₄₄ and Ser₂₁₄₅ on protein CG3523 that are phosphorylated by PKC and shaggy, respectively. These examples illustrate the importance of identifying the precise phosphorylation sites among

neighboring phosphorylatable residues as they can be modified by different kinases and/or modulated differentially upon cell stimulation. In addition, the capability to identify phosphopeptides isomers provides an interesting approach to determine densely phosphorylated regions that can act as platforms to mediate intramolecular or intermolecular electrostatic interactions.

While the scope of the present study was focused on phosphopeptides isomers, the application of these algorithms can be expanded beyond phosphoproteomics to determine the occurrence and separation of different post-translational modifications giving rise to isomeric congeners such as deamidation, *N*- or *O*-glycosylation.

4.6. Conclusion

This investigation presents the first detailed study on phosphopeptide isomers and describes two algorithms to facilitate their detection in complex cell extracts. This led to two unexpected findings about the occurrence and types of phosphopeptide isomers present in large-scale phosphoproteomics studies. First, the proportion of phosphopeptide isomers typically represents 3-6 % of all phosphopeptides identified from TiO₂-enriched digests of mouse, rat or fly protein extracts. While monophosphorylated peptides were clearly more abundant in these datasets, it is possible that this proportion is underestimated in view of the preferential selection of TiO₂ affinity media, and the reduced MS sensitivity of multiply phosphorylated peptides. Phosphorylated sites that are closely located to each other could act cooperatively to modulate protein activities via intramolecular or intermolecular electrostatic interactions. Second, the location of the phosphorylated residue within the same peptide sequence can significantly affect its physicochemical properties leading to conformational change and variable retention time on reverse phase column. The distribution of shift in retention time of more than 110 isomeric pairs indicated a variation of 0.5-7.0 min, where more than 56% were observed within 2 min of each other. Closely eluting phosphopeptide isomers are often difficult to detect in complex biological extracts due to ambiguity in the site localization and the stochastic nature of data dependant MS/MS

acquisition. Poor fragmentation of phosphopeptides still represent a major challenge for the confirmation of positional isomers, and the number of isomers reported here should be considered as a minimum boundary of occurrence. Unambiguous localization of phosphorylation sites requires the presence of sequence-specific fragment ions that can still remain elusive even when using different activation modes, as evidenced in the present study. More efficient fragmentation methods for phosphopeptides will further expand our capability to distinguish closely related positional isomers.

Targeted LC-MS/MS using inclusion list yielded complementary information on phosphopeptide isomers, although site localization could not be obtained for 25% of targeted precursor ions irrespective of the activation mode used (e.g. HCD, ETD or CID). The use of inclusion list with repeated MS/MS scans enabled the resolution of phosphopeptide isomers that are either missed or incorrectly assigned due to non optimal acquisition of MS/MS spectra using DDA.

Peptide separation with on-line 2D-LC with SCX and C₁₈ stationary phases was used to reduce sample complexity and to determine the occurrence of phosphoisomers in a large dataset and. Several alternative methods such as ion mobility spectrometry techniques (e.g. FAIMS) could be used to separate co-eluting isomeric phosphopeptides as shown previously[1]. However, additional modifications to the current algorithm would be required to fully integrate compensation voltage scans for isomer detection. For convenience, the present algorithm was developed for the detection of isomers separated by conventional chromatography techniques.

Our approach can also unravel confounding non phosphorylated peptides with closely related mass, and phosphopeptides with alternate conformations or cleavage sites that typically represented 7% of suspected phosphopeptide isomers. Importantly, the data mining approach developed here for the identification of phosphopeptide isomers extend beyond a simple cataloguing exercise, and can be of practical application for the confirmation of phosphorylation sites for subsequent mutagenesis experiments, and for the identification of other types of peptide isomers or conjugates.

4.7. Acknowledgments

We thank Éric Bonneil for technical assistance and Dev Sriranganadane for valuable comments. MC acknowledges scholarships from the Fonds de recherche sur la nature et les technologies du Québec (FQRNT) and the Faculté des études supérieures et postdoctorales (MC and GB). IRIC is supported in part by the Canadian Center of Excellence in Commercialization and Research, the Canada Foundation for Innovation, and the Fonds de recherche du Québec – Santé (FRQS). This work was carried out with the financial support of operating grants from the National Science and Engineering Research Council (NSERC) and from the Canada Research Chair program.

Chapitre 5 : Discussion, Conclusion et Perspectives

5.1. Discussion

La phosphorylation des protéines est une modification post-traductionnelle dynamique et réversible qui induit des changements structuraux au niveau de ses substrats. Son étude est capitale afin d'appréhender les mécanismes impliqués dans diverses voies de signalisation. La phosphoprotéomique est à ce jour la méthode la plus performante pour l'étude exhaustive des événements de phosphorylation. Il est en effet possible d'identifier et de quantifier des milliers d'évènements de phosphorylation en plus de pouvoir localiser avec précision le site modifié. Cette stratégie haut-débit permet de détecter avec une grande sensibilité (de l'ordre du fmol) la phosphorylation d'un résidu tout en permettant une quantification précise sur une large gamme linéaire dynamique ($\geq 5\ 000$ pour l'Orbitrap Velos).

5.1.1. Les défis en phosphoprotéomique

L'identification et la quantification des événements de phosphorylation protéiques à grande échelle ainsi que la détermination de leur localisation exacte représentent un défi considérable. Des avancées importantes ont permis de contourner certaines difficultés telles que la suppression d'ionisation des phosphopeptides, la perte de neutre lors de la fragmentation, la faible abondance des formes modifiées ou encore la présence de phosphoisomères. Ces obstacles peuvent réduire considérablement la chance d'identification des phosphopeptides par spectrométrie de masse. Des améliorations remarquables ont été réalisées au cours des dix dernières années afin de faciliter l'identification de phosphoprotéines. En effet, l'utilisation de techniques de chromatographie d'affinité telles que IMAC[164] ou TiO_2 [173] afin d'enrichir un mélange de peptides en phosphopeptides s'est avérée être une méthode très efficace. De plus, les spectromètres de masse sont de plus en plus sensibles et permettent de quantifier des changements d'intensité ionique de manière fiable sur une large gamme linéaire dynamique. L'utilisation de modes de fragmentation complémentaires tels que CID, ETD et

HCD permet également d'augmenter le pourcentage de couverture lors de l'identification de phosphoprotéines. La mise en place de méthodes de séparation par chromatographie liquide en une ou deux dimensions (MudPIT) avant l'analyse par MS permet de réduire la complexité de l'échantillon injecté et ainsi maximiser le nombre d'identifications. Cependant, on estime que seulement un tiers des peptides introduits dans le spectromètre de masse est séquencé par MS/MS à cause de la trop grande complexité de l'échantillon. Dans cette thèse nous avons décidé d'introduire une étape de séparation supplémentaire qui est le FAIMS.

5.1.2. L'utilisation de LC-FAIMS-MS/MS en phosphoprotéomique

Dans le **Chapitre 2**, nous avons évalué les bénéfices analytiques de l'utilisation du FAIMS couplé à la chromatographie liquide et à la spectrométrie de masse lors d'études phosphoprotéomiques. Dans ce contexte, la spectrométrie par mobilité ionique constitue une approche complémentaire en séparant les ions en phase gazeuse selon leur mobilité ionique différentielle dans un champ électrique élevé[213]. Le FAIMS est une méthode de séparation orthogonale à la LC-MS.

Afin de comparer de manière rigoureuse l'approche FAIMS par rapport à une étude par électronébulisation régulière, nous avons analysé des phosphopeptides purifiés à partir de cellules S2 de drosophile en triplicatas biologiques dans un mode 2D-LC sur un LTQ-Orbitrap XL (**Figure 2.1**). L'approche sans FAIMS a été réalisée de manière conventionnelle. Pour l'étude FAIMS, les identifications peptidiques ont été effectuées à partir de 5 injections distinctes à 5 CVs différents (-42 V, -38 V, -34 V, -30 V et -24 V) tandis que les données d'intensités ont été recueillies lors d'une injection balayant les 5 CVs de manière consécutive. Nous avons démontré une augmentation dans la détection de phosphopeptides de 19 % lors de l'utilisation du FAIMS comparativement à l'étude sans FAIMS ce qui est cohérent avec la littérature[233]. La reproductibilité des analyses était très similaire avec des RSD de 45 % et 42 % pour les études avec FAIMS et sans FAIMS, respectivement, ce qui prouve la robustesse du système. La fragmentation des peptides a été

réalisée dans un mode d'arbre de décision CID/ETD où la méthode la plus adéquate était choisie en temps réel selon la charge et le m/z du peptide. Nous avons identifié 3 476 phosphopeptides lors de l'expérience FAIMS et 2 168 lors de l'étude non-FAIMS pour un total de 4 445 phosphopeptides uniques (67 % d'entre eux sont identifiés par CID et le reste par ETD) (**Figure 2.2**). Nous avons démontré la complémentarité de ETD vis-à-vis de CID pour la fragmentation des phosphopeptides ce qui est en accord avec les conclusions de Swaney, *et al.*[247]. Environ un tiers des spectres MS/MS assignés a été identifié avec ou sans FAIMS tandis que 51 % d'entre eux sont uniques à FAIMS. Environ un quart est uniquement identifié sans FAIMS. Un chevauchement complet est impossible car le fait d'utiliser des sauts de CVs en FAIMS peut compromettre la détection de certains peptides pour lesquels la transmission n'est pas optimale pour ces CVs fixés de manière arbitraire. Un gain considérable dans le nombre d'identifications de phosphopeptides de plus faible intensité (inférieure à 300 000 comptes) a été obtenu avec FAIMS avec 1 188 phosphopeptides identifiés dans cette catégorie contre 394 seulement sans FAIMS. Cette technique nous a permis d'augmenter notre limite de détection et d'identifier des ions de plus faible abondance en réduisant la transmission de bruit de fond (ions mono-chargés contaminants) vers le spectromètre de masse. L'augmentation dans le nombre de peptides identifiés est aussi due à une meilleure séparation de peptides qui co-éluent en LC et qui sont désormais séparables par FAIMS (**Figure 2.3**). Bien que sa résolution soit plutôt modeste (~ 15), ce dispositif permet une séparation de phosphoisomères non distinguables par LC et qui sont souvent mal identifiés à cause de la génération de spectres MS/MS mixtes. FAIMS peut séparer ces phosphopeptides isomériques pour lesquels le profil de transmission de CVs est décalé, ce qui permet alors l'acquisition de spectres MS/MS distincts (**Figure 2.4**). Le FAIMS est sans conteste une méthode de choix lors d'études phosphoprotéomiques.

5.1.3. La voie de signalisation de l'insuline

La voie de signalisation de l'insuline est très conservée chez la drosophile par rapport aux mammifères et intervient dans le métabolisme des lipides et des carbohydrates ainsi que dans la croissance et la prolifération cellulaire[48, 49]. La stimulation à insuline induit l'activation d'une cascade de phosphorylation en aval de son récepteur InR de type RTK et fait intervenir notamment les protéines Pi3K, AKT, ERK et TOR qui sont bien connues dans la littérature. Comprendre les événements de phosphorylation impliqués suite à l'activation des RTKs est d'une importance capitale car plusieurs d'entre eux sont activés de manière anormale dans différents types de cancers[29]. Il est alors d'intérêt clinique d'étudier cette voie par une approche phosphoprotéomique afin d'avoir une vue globale et approfondie des mécanismes mis en jeu. Ces études peuvent permettre le développement de nouveaux traitements thérapeutiques ciblés.

5.1.4. L'utilisation de FAIMS pour déterminer les événements de phosphorylation associés à la voie de l'insuline chez la drosophile

Nous avons appliqué la stratégie LC-FAIMS-MS/MS afin de disséquer les événements de phosphorylation intervenant suite à la stimulation de cellules S2 par l'insuline. Notre but était de comparer les phosphoprotéomes différentiels de cellules contrôles et traitées à l'insuline afin d'identifier et quantifier des changements de phosphorylation.

Suite au traitement à l'insuline pendant 8 min à une concentration de 0.86 μM , nous avons détecté par immunobuvardage de type Western l'activation simultanée des voies MAPK et PI3K/AKT (**Figure 2.5. A**) qui sont des contrôles positifs attendus. Ensuite, l'étude phosphoprotéomique globale comparant les phosphoprotéomes de cellules contrôles et traitées à l'insuline a mis en évidence 32 phosphopeptides différentiellement exprimés (18 sous-phosphorylés et 14 sur-phosphorylés) sur les 2 660 phosphopeptides identifiés et quantifiés (soit environ 1 %). Fait intéressant, nous avons confirmé l'activation de MAPK (ou rolled, homologue de ERK) (T₂₄₈/Y₂₅₀) (**Figure 2.5. B**) ainsi que celle de Pi3K (S₁₃₆)

(**Figure 2.7. A**) par spectrométrie de masse en démontrant l'augmentation de leur phosphorylation suite à la stimulation. Cette étude à grande échelle a montré la modulation de phosphorylation de protéines impliquées dans la traduction et transcription (nucleosome assembly protein 1, T₁₂₀), dans le trafic vésiculaire (cystein string protein, S₁₂), l'assemblage du cytosquelette (kinesin heavy chain, S₇₂₄) et le métabolisme des lipides (lipin homologue, S₂₂₉₁) (**Table 2-I**). La stimulation à l'insuline semble affecter plusieurs processus cellulaires et plusieurs ont déjà été rapportés dans la littérature. C'est le cas notamment du rôle de l'actine pour la propagation des signaux médiés par l'insuline[400] ou encore de l'importance du transport de vésicules dans la régulation de l'homéostasie du glucose dans la circulation sanguine[401]. Nous pensons que la phosphorylation de sites précis sur ces cibles d'intérêt joue un rôle déterminant pour plusieurs mécanismes physiologiques. Nous avons ensuite conduit une analyse de motifs grâce à l'outil Motif-X[9] afin de déterminer les séquences consensus entourant les sites de phosphorylation identifiés dans nos données et avons découvert que près de 50 % des protéines dont la phosphorylation est modulée par l'insuline sont phosphorylées sur un motif SXXXD pouvant correspondre à une cible de la protéine kinase CK2 (**Figure 2.8**). Cette exploration nous a permis d'émettre l'hypothèse que CK2 pourrait jouer un rôle clé dans la régulation de la voie de l'insuline. C'est ce que nous avons voulu vérifier dans le chapitre suivant.

5.1.5. Le rôle de la protéine kinase CK2 dans la voie de l'insuline

Suite aux découvertes décrites précédemment concernant de nouveaux interacteurs de la voie de l'insuline, nous avons cherché à mieux comprendre le rôle de CK2. Le **Chapitre 3** avait pour objectif l'identification des événements de phosphorylation dépendant à la fois de la stimulation à l'insuline et de CK2 dans les cellules S2. CK2 est une kinase exprimée de manière ubiquitaire dans la cellule et qui est impliquée dans de nombreux processus tels que la transcription et la traduction, la croissance ou encore la prolifération cellulaire. Elle est décrite comme étant active de manière constitutionnelle mais les mécanismes qui

modulent son activité restent encore mal compris[124]. Son activation suite à la stimulation à l'insuline demeure un sujet de discussion dans la littérature[129].

Nous avons utilisé une approche combinant l'utilisation de RNAi afin d'inhiber l'expression de CK2 et une stratégie 1D-LC-FAIMS-MS/MS sur un Orbitrap Velos en mode HCD. L'inhibition de l'expression de CK2 a été vérifiée par immunobuvardage de type Western ainsi que par RT-q-PCR et a été estimée à environ 80 % (**Figure 3.2**). Nous avons ensuite comparé les phosphoprotéomes de cellules contrôles, stimulées à l'insuline et/ou traitées avec un RNAi CK2. Au total, nous avons identifié 2 067 phosphopeptides uniques parmi lesquels 51 voient leur intensité modulée suite au traitement à l'insuline (18 sous-phosphorylés et 33 sur-phosphorylés) (**Figure 3.3. A**). Nous avons identifié les protéines Pi3k92E (Y₁₃₈) et Chico (S₅₀₂) parmi les protéines sur-phosphorylées qui témoignent de la bonne transduction du signal suite à l'insuline. Par la suite, une combinaison d'outils de prédiction de kinase (PPSP[353] et NetworKIN[294]) et de reconnaissance de motifs (Motif-X) a été utilisée afin de déterminer les kinases impliquées dans la phosphorylation des sites identifiés. Encore une fois, les logiciels s'accordent pour démontrer que la proportion de sites potentiellement phosphorylables par CK2 est significativement plus grande pour les sites modulés par l'insuline comparativement au jeu de données global (**Figure 3.3. B**), ce qui conforte les observations décrites au **Chapitre 2**. CK2 est la seule kinase qui démontre un tel niveau d'enrichissement suite à la stimulation à l'insuline ce qui en fait donc une protéine d'intérêt.

En combinaison avec le traitement à l'insuline, les cellules S2 ont été traitées ou non avec un RNAi pour CK2 et 148 cibles ont été découvertes comme étant différemment phosphorylées suite à l'inhibition de CK2 (65 sur-phosphorylées et 83 sous-phosphorylées) (**Figure 3.4. A**). Des contrôles positifs tels qu'Eif4b (S₄₆) et Top2a (S₁₃₄₄) ont été mis en évidence parmi les 83 comme étant des cibles directes connues de CK2. De nouvelles cibles ont également été identifiées telles que la protéine Hu-li tai shao (T₅₈₆, S₆₃₀ et T₆₃₄) qui est l'homologue de l'adducine chez le mammifère. Au total, nous avons identifié 6 phosphoprotéines (CG30085 [S₉₅₉], su(var)205 [S₁₅], scny [S₆₅₀], protein CDV3 homolog [S₁₃₄], D1 [S₃₀] et mu2 [S₆₃₀]) pour lesquelles l'intensité était positivement modulée suite à

la stimulation à l'insuline puis négativement régulée après le traitement par RNAi CK2. La phosphorylation de ces protéines est dépendante de l'insuline et de CK2 (de manière directe ou indirecte). Afin d'identifier les cibles directes de CK2, nous avons conduit des essais kinases *in vitro* grâce à l'utilisation de peptides synthétiques. Un ensemble de 134 peptides dont les séquences étaient basées sur les identifications précédentes dans les cellules S2 a été synthétisé. Nous avons réalisé une étude cinétique afin de regarder l'évolution de la phosphorylation de chaque peptide suite à sa réaction avec la protéine CK2 recombinante dans un seul et même tube afin que les peptides restent en compétition. La même expérience a été réalisée avec CK1 comme contrôle négatif étant donné que les deux kinases ont pour cible des séquences consensus constituées d'acides aminés acides. Par spectrométrie de masse, nous avons observé une phosphorylation sur 29 peptides synthétiques identifiés comme des cibles directes de CK2 et chacun d'entre eux contenait la séquence consensus pour CK2 (**Figure 3.6**). Ces résultats prouvent la spécificité de l'essai et la fiabilité des observations. Sur ces 29 peptides, 15 d'entre eux avaient été identifiés comme sous-phosphorylés suite au RNAi CK2. La protéine heterochromatin protein 1 (S₁₅) fait partie de ces 15 cibles directes de CK2 et est déjà connue dans la littérature. Nous avons montré que sa phosphorylation est en plus dépendante de l'insuline. Bien que les essais *in vitro* réalisés sur des peptides ne tiennent pas compte de la structure tridimensionnelle des protéines, ils permettent néanmoins de cribler une centaine de cibles de manière simultanée.

Finalement, nous avons voulu vérifier si CK2 était plus active suite à la stimulation à l'insuline. Pour y parvenir, nous avons conduit un essai kinase *in vitro* sur le peptide substrat de CK2 (CK2SP) en réaction avec un extrait protéique provenant de la lyse de cellules S2 contrôles ou stimulées à l'insuline. Nous avons démontré que CK2 est active de manière constitutionnelle dans les cellules contrôles mais n'est pas plus active suite à l'insuline bien qu'on observe des changements positifs sur des substrats potentiels ayant le site consensus approprié pour CK2 (**Figure A3. 5**). Ainsi, un autre mécanisme est probablement impliqué afin de moduler la phosphorylation de substrats de CK2. Nous émettons l'hypothèse qu'un changement conformationnel en réponse à l'insuline puisse

rendre certains substrats plus susceptibles d'être phosphorylés par CK2. Il est également possible qu'une protéine phosphatase partageant le même motif consensus que CK2 intervienne simultanément afin de balancer l'effet de CK2 sur ses substrats. L'activité de CK2 peut également dépendre de sa localisation cellulaire et mener à la phosphorylation différentielle de ses substrats selon le compartiment où elle est exprimée préférentiellement.

D'une manière générale, peu d'information est disponible concernant la régulation de CK2 par des facteurs externes. Plusieurs hypothèses peuvent être avancées quant au rôle de CK2 dans la voie de signalisation de l'insuline. Il a déjà été décrit que les niveaux de CK2 ainsi que son activité kinase sont plus élevés dans les tumeurs solides ainsi que dans les tissus prolifératifs[402]. De plus CK2 est connue pour favoriser la croissance cellulaire et réguler négativement la mort cellulaire programmée. Guerra, *et al.* ont montré que les sous-unités de CK2 pouvaient former un hétérocomplexe stable avec Akt *in vivo* et augmenter l'activité de Akt de manière synergique[403]. Cependant, ils ont également démontré que la phosphorylation de Akt induite suite à la stimulation avec IGF-1 ne modulait pas l'interaction de Akt avec CK2. Il faut noter que l'effet de l'insuline n'a pas été testé. Il a été rapporté que CK2 interagit avec de nombreuses protéines ce qui peut avoir diverses conséquences en terme d'activation ou d'inhibition dépendamment de la protéine en question[124]. Par exemple, l'interaction de CK2 β avec c-Mos inhibe l'activité kinase de c-Mos[404] tandis que la formation d'un complexe CK2 β -Chk1 va augmenter l'activité de Chk1[405]. Il est alors envisageable que, suite au traitement à l'insuline, CK2 forme un complexe avec d'autres protéines *in vivo*, ce qui pourrait moduler leurs activités dans des conditions particulières. L'identification de protéines formant un complexe avec CK2 et leur modulation pourrait expliquer en partie la régulation de substrats de CK2.

Le rôle de CK2 dans la production d'insuline par des cellules pancréatiques de type β a déjà été démontré suite à l'identification du facteur de transcription Pdx-1 (T₂₃₁, S₂₃₂) comme substrat de CK2[130, 406]. En effet, Pdx-1 permet de réguler la production d'insuline et de Glut-2 tout en régulant l'expression de la glucokinase, ce qui mène à la modulation de l'entrée de glucose. Ils ont émis l'hypothèse d'une régulation négative de la

sécrétion d'insuline via la phosphorylation de Pdx-1 par CK2. L'hypothèse de boucle de rétroaction négative suite à la phosphorylation de protéines cibles est également plausible.

Il a déjà été démontré que la phosphorylation de PTEN par CK2 (S₃₇₀, S₃₈₅) et par GSK3- β (S₃₆₂, T₃₆₆) semble être modulée par un traitement à l'IGF[407]. Ainsi, CK2 participerait à la modulation de l'activité phosphatase de PTEN et régulerait le métabolisme des lipides. L'effet de l'insuline sur ce mécanisme n'a pas été testé mais l'hypothèse de la régulation de diverses phosphatases par CK2 est envisageable.

Lors de notre étude, certaines protéines reliées au dommage à l'ADN et à la liaison à la chromatine ont démontré une modulation positive de leur phosphorylation suite au traitement à l'insuline puis une déphosphorylation suite au RNAi CK2. Des études ont déjà rapporté un lien entre la voie de signalisation de l'insuline et des phénomènes de dommage à l'ADN locaux liés à la réponse du système immunitaire inné afin de réguler la croissance et l'activité métabolique chez les larves de drosophile[367]. Il a été décrit que la répression de la voie de signalisation de l'insuline ou de l'IGF induit la voie NF κ B/Relish dans le corps gras, ce qui est requis lors de la réponse aux dommages à l'ADN liés, par exemple, aux rayons UV (ultraviolets). Ils suggèrent que lorsque le dommage à l'ADN ne peut être réparé, le signal devient alors permanent et prolongé, ce qui implique que la voie de l'insuline est modulée afin d'empêcher une réponse immunitaire excessive et équilibrer la croissance cellulaire avec le processus de réparation.

CK2 est connue comme étant impliquée dans la régulation d'espèces réactives à l'oxygène et dans la formation de coupures simple ou double brins d'ADN[368-370]. Plus particulièrement, chez la drosophile, CK2 est connue pour jouer un rôle prédominant en phosphorylant des protéines facilitant la réparation des dommages à l'ADN. Par exemple, l'exposition aux UV augmente l'abondance d'un complexe composé notamment de CK2, hSpt16 et SSRP1 ce qui induit la phosphorylation de P53 (S₃₉₂)[408]. Il a été démontré également que CK2 favorise la régulation de NF κ B suite aux UV[409]. CK2 phosphoryle XRCC1 impliquée dans la réparation de coupures d'ADN simple brin ainsi que SSRP1 connue pour son rôle dans la transcription et la réplication via le remodelage de la chromatine[410]. Ces évidences semblent être en accord avec les conclusions décrites lors

de notre étude et suggèrent un lien entre la voie de signalisation de l'insuline, CK2 et le dommage à l'ADN.

Tel que mentionné précédemment, une seule phosphoprotéine (heterochromatin protein 1, S₁₅) a été identifiée dans notre étude comme modulée par l'insuline et directement phosphorylée par CK2 comme le confirme l'expérience par RNAi et l'essai kinase *in vitro*. Il faut mentionner que certaines limitations dans notre étude peuvent expliquer pourquoi d'autres cibles ont pu être manquées. Tout d'abord, il faut noter que l'inhibition par RNAi CK2 n'est pas complète et que certains substrats directs de CK2 n'ont pas démontré de modulation d'intensité suite au RNAi car même une faible proportion de CK2 active suffisait à leur phosphorylation. Par ailleurs, certaines protéines dont la phosphorylation est négativement modulée par le RNAi CK2 nécessitent peut-être une conformation tridimensionnelle particulière pour leur phosphorylation et n'ont pu être validées lors de l'essai *in vitro* utilisant des peptides synthétiques dépourvus de structure.

5.1.6. Comparaison des données phosphoprotéomiques avec la littérature disponible sur la voie de l'insuline

Tel que décrit dans la **section 1.5.3**, d'autres études phosphoprotéomiques visant l'étude de la voie de l'insuline ont déjà été conduites et décrites dans la littérature. Cependant, peu d'information à grande échelle est disponible pour la drosophile comparativement aux mammifères.

En 2006, Schmelzle *et al.* ont réalisé une étude cinétique de la voie de signalisation de l'insuline dans des adipocytes 3T3-L1 avec des temps d'incubation de 5, 15 et 45 minutes[80]. Grâce à une étape d'immunoprécipitation sur des phosphotyrosines puis un enrichissement par IMAC, ils ont identifié 122 sites dont 89 sont différenciellement phosphorylés suite à un traitement à l'insuline (quantification par i-TRAQ, changement supérieur à 1.3 fois). Des triplicatas biologiques ont été analysés et ont mené notamment à l'identification de ERK-1 (T₂₀₃, Y₂₀₅), ERK-2 (T₁₈₃, Y₁₈₅), Inr (T₁₃₄₄, Y₁₃₄₅, Y₁₃₅₁) et Irs1 (Y₁₃₇₁) comme étant des cibles sur-phosphorylées avec un changement d'intensité supérieur

à 10 fois après 5 minutes d'incubation. Ils ont par ailleurs mis en évidence des protéines impliquées dans le transport du glucose et la translocation de GLUT4 telles que Syntaxin 4 (Y₂₅₁, Y₁₁₅), Munc18c (Y₁₅₂₁), EH-domain containing 2 protein (Y₄₅₃) et Annexin 2 (Y₂₀, Y₂₃₈). En 2008, Krüger, *et al.* ont publié une étude de la voie de signalisation de l'insuline dans des cellules d'adipocytes bruns[81]. La méthode impliquait également une étape d'immunoprécipitation des phosphotyrosines suivie d'une quantification par SILAC et a permis l'identification de 40 sites différenciellement phosphorylés suite au traitement à l'insuline pendant 1, 5, 10 ou 20 minutes. Un changement d'intensité supérieur à 1.3 était considéré comme significatif et la phosphorylation différentielle de InR (Y₁₁₈₅), Irs1 (Y₈₉₁, Y₉₈₃), Irs2 (Y₉₇₀), Aps (Y₆₁₈) et Gab-1 (Y₂₅₉) a entre autres a été démontrée. Des interacteurs de la voie MAPK (tels que ERK-2 [Y₁₈₅]) ont aussi été identifiés avec un maximum d'activation après 5 minutes. De manière intéressante, l'hypothèse d'un lien entre l'insuline et la voie de signalisation du calcium a été émise suite à l'identification de PDZK11 (Y₇) comme modulé par l'insuline. Des immunobuvardages de type Western à l'aide d'anticorps anti-pY ont été réalisés comme première étape de validation.

Lors des deux études décrites précédemment, bien que l'utilisation d'une étape d'immunoprécipitation anti-pY restreigne considérablement le nombre de sites identifiés, le fait d'étudier la voie de signalisation de l'insuline de manière cinétique ajoute un aspect dynamique essentiel à leurs études, ce qui n'a pas été exploré totalement dans cette thèse. En effet, plusieurs temps d'incubation relativement courts ont été testés par immunobuvardage de type Western (4, 8, 30 et 60 min) (**Figure 2.5**) lors du traitement des cellules S2 à l'insuline afin de s'assurer que le niveau d'expression protéique ne soit pas affecté et que seuls les changements de phosphorylation soient mesurés. Cependant, pour l'analyse phosphoprotéomique, nous avons ensuite choisi un temps d'incubation fixe de 8 min afin d'observer simultanément les effets transitoires de l'activation de la voie MAPK et les effets prolongés de la voie Pi3K/Akt.

En 2010, Larance, *et al.* ont étudié l'effet de l'insuline sur la phosphorylation de protéines se liant à la protéine 14-3-3 pour se concentrer exclusivement sur l'étude des substrats de Akt dans des cellules myotubes L6 de souris[82]. Suite à une étape

d'immunoprécipitation grâce à un anticorps dirigé contre 14-3-3, ils ont identifié et quantifié par SILAC 400 protéines dont 28 (telles que FOXO3A, PRAS40, GAB2, BRF1 et TSC2) démontraient une liaison différentielle à 14-3-3 et dont plusieurs possédaient un site consensus pour Akt. Un lien entre la régulation post-transcriptionnelle de l'ARNm (via la phosphorylation de EDC3 [S₁₆₁] par Akt) et la voie de signalisation de l'insuline a été démontré. Il ne s'agit par contre pas d'une étude phosphoprotéomique globale mais plutôt une analyse ciblée qui a pour limite la spécificité de l'anticorps utilisé.

Simultanément aux études décrites dans cette thèse, Monetti *et al.* ont publié en 2011 une étude exhaustive de la voie de signalisation de l'insuline dans le foie de souris *in vivo* grâce à une expérience phosphoprotéomique utilisant SILAC et une méthodologie 'spike-in' impliquant des standards marqués[84]. Ils ont identifié 20 000 sites de phosphorylation (15 000 avec une confiance de localisation supérieure à 75%) dont 10 000 d'entre eux ont été quantifiés. Cette étude *in vivo* permet de recueillir de l'information complémentaire à celle disponible lors de l'utilisation de modèles en cultures cellulaires. En effet, la contribution des différents tissus environnants permet de mesurer l'influence globale et de prendre en considération différents facteurs pouvant influencer les effets en réponse au traitement à l'insuline. La robustesse et la reproductibilité de l'étude ont été démontrées en triplicatas techniques à partir de deux souris pour chaque condition. Tout comme nous, ils ont identifié une faible proportion de sites (1 000 sites) modulés par l'insuline avec un changement d'intensité supérieur à 2. De manière intéressante, leur étude a permis d'identifier plusieurs sites de phosphorylation sur des médiateurs de la voie dont Irs1 (S₃₀₇ et S₃₀₂) et Irs2 (T₅₁₇). Concernant la voie Ras-MAPK, aucun des sites identifiés ne montrait de changement d'intensité suite au traitement à l'insuline contrairement à notre étude. En effet, les souris ont été traitées pendant 15 min à l'insuline ce qui est certainement trop long pour pouvoir observer les effets transitoires dans cette voie. Ils ont identifié des membres de la voie Pi3K/Akt (S₄₇₃ de Akt) ainsi que mTor mais peu d'entre eux montraient des changements significatifs de phosphorylation. Bien que la technique utilisée soit remarquable, aucune validation biologique n'a été effectuée contrairement à ce qui a été décrit dans cette thèse. Par contre, la technique de quantification semble très robuste bien

que les outils de validation utilisés pour rapporter les données quantitatives sont peu décrits. Cela représente cependant un des aspects principaux à prendre en considération lors de la description d'études globales.

Chez la drosophile, à ma connaissance, aucune étude phosphoprotéomique globale n'a été rapportée dans la littérature afin d'étudier la voie de signalisation de l'insuline. Cependant, des études ciblées ont utilisé la phosphoprotéomique afin d'étudier précisément certains complexes. C'est le cas d'une étude publiée par Przybylski *et al.* en 2010 dont l'intérêt est porté sur les complexes protéiques formés autour de la protéine Chico (homologue de Irs) par i-TRAQ suite à un traitement à l'insuline dans des cellules Kc167[83]. Une étape d'immunoprécipitation a été effectuée afin d'identifier les protéines du complexe (14-3-3, heat shock protein, actine...) puis un traitement à la phosphatase a été utilisé pour identifier les sites de phosphorylation sur Chico (S₂₅₅, S₂₅₈, S₂₆₀, S₂₈₆, S₂₈₇, S₃₀₀, S₃₁₉, S₃₄₂, S₄₇₁, S₆₂₀, S₈₁₃, S₉₀₅). D'autre part, Radimerski *et al.* se sont intéressés à la phosphorylation multiple de la protéine ribosomale S6 en réponse à l'insuline dans des cellules Kc167[411]. Bien que la spectrométrie de masse ait été utilisée pour déterminer l'identité de la protéine phosphorylée, les résidus S₂₃₃, S₂₃₅, S₂₃₉, S₂₄₂ et S₂₄₅ ont été identifiés par dégradation d'Edman.

De manière générale, les études décrites chez la drosophile ne couvrent pas l'aspect global et quantitatif du phosphoprotéome et donc ne permettent pas d'évaluer la régulation de sites spécifiques selon différentes conditions. Nous avons publié une étude qualitative et quantitative complète pouvant servir de référence en vue d'études ciblées visant l'élucidation de la fonction biologique de sites de phosphorylation précis régulés par l'insuline (ce qui n'avait jamais été décrit auparavant). Il faut noter que certains événements de phosphorylation bien connus dans la voie de l'insuline n'ont pas été observés lors de notre étude pour différentes raisons. D'abord, la phosphorylation est une réaction extrêmement rapide qui peut se produire en l'espace de quelques secondes. Ainsi, certains événements se produisant en amont de la voie suite à la liaison de l'insuline sur son récepteur peuvent avoir été manqués lors de notre analyse (8 minutes de traitement). De plus, il faut garder à l'esprit que les étapes de collection et de lavage des cellules (bien

qu'effectuées sur la glace) prennent du temps et que les réactions enzymatiques ne sont pas stoppées instantanément. Par ailleurs, la méthode de lyse cellulaire utilisée ne permettait pas d'extraire les protéines membranaires, ce qui explique possiblement pourquoi les événements de phosphorylation intervenant sur le récepteur par exemple n'ont pas été observés. Il faut mentionner également que certains résidus sont phosphorylés sur des peptides dont les caractéristiques physicochimiques défavorisent leur identification par LC-MS/MS. En effet, certains phosphopeptides sont trop courts, trop longs, peu ionisables ou encore trop hydrophiles et ne seront pas détectés.

5.1.7. La détection de phosphoisomères lors d'études phosphoprotéomiques

Lors de l'analyse des données phosphoprotéomiques conduites dans les cellules S2 de drosophile, nous avons mis en évidence la présence de phosphoisomères qui co-éluent en LC mais qui sont séparables par FAIMS. C'est le cas notamment des isomères DKSEGDpSPNIETPKK et DKSEGDSPNIEpTPKK appartenant à la protéine small-minded (CG8571). Les phosphopeptides isomériques ont le même m/z et sont souvent confondus en LC car ils sont constitués de la même séquence peptidique et ont les mêmes propriétés physico-chimiques. Nous nous sommes alors interrogés sur la fréquence de ces cas car cet aspect est peu rapporté dans la littérature. Le fait de pouvoir localiser avec confiance un site de phosphorylation sur un peptide contenant plusieurs résidus phosphorylables est essentiel afin d'y accorder une fonction biologique et mener des expériences de validation. Environ un tiers des sites de phosphorylation identifiés par une approche protéomique a une localisation ambiguë et des outils tels que A-score[272] permettent d'attribuer un score reflétant la confiance dans la position du phosphate.

Nous nous sommes intéressés à cette problématique dans le **Chapitre 4** et avons décrit l'implémentation de deux algorithmes complémentaires pour détecter des phosphopeptides isomériques lors d'analyses phosphoprotéomiques à grande échelle. Le premier script permet de cibler les isomères distinguables par LC-MS en se basant sur leurs

différences de temps de rétention. Il a été utilisé afin d'évaluer la proportion de phosphoisomères dans des cellules J774 de souris et IEC-6 de rat (4 798 et 15 700 phosphopeptides uniques identifiés, respectivement). Nous avons déterminé que moins de 1 % des phosphopeptides identifiés constituent des phosphoisomères séparables par LC. Ce nombre ne tient donc pas compte des isomères qui co-éluent ou qui n'ont pas été séquencés par MS/MS à cause de la fenêtre d'exclusion dynamique. Nous avons identifié 64 phosphoisomères (**Table A4- II**) dans ces deux espèces et nous nous sommes intéressés à leurs propriétés physico-chimiques afin de déterminer une tendance expliquant leur décalage de temps de rétention en LC. Aucune prédisposition concrète ne permettait d'expliquer ce phénomène que ce soit par un calcul d'hydrophobicité locale ou encore la présence d'interactions électrostatiques. Par contre, nous avons remarqué que dans trois quarts des cas, le phosphopeptide dont le résidu modifié est le plus éloigné de l'extrémité N-ter élue en premier (**Figure A4. 2**). Ensuite, en utilisant le logiciel de modélisation 3D PEP-FOLD, nous avons mis en évidence une différence de conformation entre les peptides VGGpSSVDLHR et VGGSpSVDLHR de la delta-caténine qui pourrait expliquer le décalage de temps de rétention de 1.4 min entre les deux formes (Erreur ! Source du renvoi introuvable.). Cependant, la majorité des cas de phosphoisomères observés ne pouvaient être expliqués par la prédiction de modèles 3D.

Nous avons ensuite testé notre algorithme sur 26 peptides synthétiques isomériques dont les séquences étaient basées sur les identifications obtenues dans les cellules S2 afin de prouver notre concept. Nous avons injecté chacun des 26 phosphopeptides mono-phosphorylés seuls puis en combinaison d'isomères et nous avons découvert que 18 d'entre eux étaient totalement séparables par LC et pour lesquels la position du site modifié était rapportée avec une bonne confiance par ETD ou CID. La plupart d'entre eux était détectée par notre algorithme mais nous avons également observé des artefacts dus à la présence de conformères en solution, ce qui a déjà été rapporté dans la littérature[392].

Nous avons mis à profit cet algorithme afin de conduire une étude ciblée chez la drosophile (10 110 sites de phosphorylation identifiés) où nous avons confirmé la présence de 77 vraies paires isomériques et 199 paires ambiguës pour lesquelles la localisation du

phosphate était incertaine. Nous avons ensuite procédé à l'analyse ciblée des candidats potentiels grâce à une liste d'inclusion construite à partir des résultats de l'algorithme et nous avons identifié 86 phosphoisomères dont 46 avaient déjà été vus lors de l'injection en DDA. Ainsi, notre approche ciblée a permis de mettre en évidence 40 nouvelles paires isomériques. La plupart des paires identifiées uniquement en DDA ont également été détectées par l'algorithme mais n'ont pas mené à une identification fiable lors de l'inclusion (cycle trop rapide, abondance...) et n'ont donc pas été retenues. L'approche utilisant l'algorithme est complémentaire au mode DDA et les deux modes combinés ont permis de détecter 117 paires d'isomères non ambigus avec environ 50 % de recoupement (**Table A4- IV**). Il faut noter aussi ici la présence d'artefacts considérés comme de potentiels phosphoisomères par l'algorithme tels que des peptides générés avec un site de clivage tryptique alternatif, des séquences peptidiques différentes ou encore des conformères (**Figure A4. 7**). Notre stratégie a aussi permis l'identification d'isomères multiplement phosphorylés tels que le peptide IPSSSSDFSK modifié sur les résidus S₃, S₄, S₅ et S₆ et identifié avec une bonne confiance (**Figure 4.6**).

Le deuxième algorithme présenté vise à détecter la présence de phosphoisomères qui co-éluent ou qui sont partiellement séparés en se basant sur leur ions fragments caractéristiques en MS/MS par CID, ETD et HCD. HCD nous permet ici de diminuer considérablement notre taux de faux positifs grâce à une meilleure exactitude de masse sur les fragments. Notre stratégie consistait à faire une première injection de l'échantillon en abaissant la fenêtre d'exclusion dynamique à 10 s afin d'obtenir un profil MS/MS le plus complet pour un ion donné. Lors d'une seconde analyse ciblée, nous avons pu mettre en évidence 23 paires de phosphoisomères partiellement séparées et confirmées par MS/MS grâce à la liste d'inclusion. Ces isomères ne sont pas identifiables lors d'une étude en DDA conventionnelle et l'utilisation d'un script a permis de les mettre en évidence de manière exclusive.

Nous avons utilisé ici une stratégie MudPIT afin de séparer les peptides par SCX puis par C₁₈ en ligne avec le spectromètre de masse. Le but du fractionnement par SCX était d'augmenter la couverture des peptides identifiés pour nos analyses statistiques et non

de séparer les phosphopeptides isomériques selon leur charge. Des approches alternatives telles que la chromatographie HILIC, sur colonne monolithique ou encore l'électrophorèse capillaire seraient envisageables et facilement exploitables par l'algorithme afin de distinguer des isomères. Il serait également pertinent de réaliser la séparation de phosphoisomères par mobilité ionique (FAIMS) et d'adapter l'algorithme en conséquence pour le traitement des données. Il serait intéressant de mener une étude exhaustive afin de tester ces différentes techniques et évaluer leur complémentarité afin d'identifier un maximum d'isomères. Il faut mentionner que le but global de notre étude était de déterminer la fréquence des isomères lors d'études à grande échelle réalisées de manière conventionnelle.

La combinaison des deux algorithmes décrits précédemment est une stratégie intéressante dans le but d'identifier avec confiance la localisation d'un site de phosphorylation précis pour ensuite effectuer par exemple des expériences de mutagenèse dirigée en vue d'étudier sa fonction. Cet outil permet également d'obtenir une quantification plus précise lorsque deux isomères sont partiellement séparés et qu'une méthode sans marquage est utilisée.

La charge négative du phosphate peut moduler la conformation de la protéine et avoir un rôle de régulation quant à son activité ou sa liaison avec d'autres molécules. Notre capacité à identifier des phosphoisomères permet de déterminer des régions à haute densité de phosphorylation qui peuvent jouer un rôle de plateforme pour établir des interactions électrostatiques. Il serait aussi envisageable d'utiliser les algorithmes décrits ici afin d'examiner d'autres modifications post-traductionnelles.

5.2. Conclusion

La phosphorylation des protéines influence leur fonction (en induisant leur activation ou désactivation), leur localisation cellulaire ainsi que leur interaction avec diverses molécules. Elle régule plusieurs processus tels que la division, la croissance et la différenciation cellulaire. La phosphorylation est réalisée suite à l'intervention de kinases et phosphatases selon un mécanisme coordonné. Une mauvaise régulation de ce réseau de signalisation suite à des mutations ou à des niveaux protéiques anormaux peut être la cause de développement de cancers ou encore de maladies métaboliques[306]. L'objectif global de cette thèse était de disséquer les événements de phosphorylation impliqués dans la voie de signalisation de l'insuline chez la drosophile. Le récepteur de l'insuline fait partie de la famille des récepteurs tyrosine kinase et son activation induit une cascade de phosphorylation très conservée impliquant des interacteurs dérégulés dans divers cas de cancers. Les résultats obtenus lors des études phosphoprotéomiques différentielles ont permis de fournir des informations pertinentes concernant des milliers d'évènements de phosphorylation intervenant dans cette voie. Ces découvertes ont été réalisables grâce à l'implémentation de techniques instrumentales à la fine pointe de la technologie. En effet, nous avons mis en évidence pour la première fois les bénéfices de l'utilisation de LC-FAIMS-MS/MS lors d'études phosphoprotéomiques à grande échelle. La séparation orthogonale des phosphopeptides par mobilité ionique en combinaison avec la LC-MS a permis d'augmenter de 50 % le nombre de phosphopeptides identifiés et a permis également la distinction de phosphoisomères. Cette technique a été appliquée pour l'étude de la voie de signalisation de l'insuline dans les cellules S2 de drosophile. Cette analyse phosphoprotéomique différentielle a démontré la modulation de phosphorylation de protéines impliquées dans la traduction et transcription (nucleosome assembly protein 1, T₁₂₀), dans le trafic vésiculaire (cystein string protein, S₁₂), l'assemblage du cytosquelette (kinesin heavy chain, S₇₂₄) et le métabolisme des lipides (lipin homologue, S₂₂₉₁). Une analyse de motifs a montré que près de 50 % des cibles régulées par l'insuline comportent un motif SXXXD consensus pour la protéine kinase CK2. Nous avons alors émis l'hypothèse que CK2 pouvait jouer un rôle dans la voie de l'insuline puis nous avons mené

une étude phosphoprotéomique combinant FAIMS-LC-MS/MS et RNAi dans les cellules S2 afin de la vérifier. Au total, nous avons identifié 6 phosphoprotéines (CG30085 [S₉₅₉], su(var)205 [S₁₅], scny [S₆₅₀], protein CDV3 homolog [S₁₃₄], D1 [S₃₀] et mu2 [S₆₃₀]) pour lesquelles l'intensité était positivement modulée suite à la stimulation à l'insuline puis négativement régulée après le traitement par RNAi CK2. La mise en place d'essais kinase *in vitro* détectés par spectrométrie de masse a permis d'analyser de manière beaucoup plus quantitative les produits de réaction phosphorylés comparativement aux méthodes conventionnelles. Cette approche a permis de valider 29 cibles directes de CK2 dont 15 corrélaient avec l'étude par RNAi. D'après nos observations, la sérine 15 de la protéine heterochromatin protein 1 est phosphorylée de manière directe par CK2 et est dépendante de l'insuline. De manière intrigante, nous n'avons pas observé d'augmentation de l'activité de CK2 suite au traitement à l'insuline des cellules S2. Ceci suggère qu'un autre mécanisme est probablement impliqué afin de moduler la phosphorylation de substrats de CK2.

Ces projets de grande envergure ont permis l'identification de milliers de sites de phosphorylation dont une faible proportion (1 %) est isomérique. C'est pourquoi nous avons utilisé une approche bioinformatique basée sur deux algorithmes complémentaires à la méthode DDA qui permet la détection efficace de phosphoisomères. Nous avons mis en évidence 64 phosphopeptides isomériques lors d'une étude combinée chez le rat et la souris et 117 cas ont été répertoriés chez la drosophile. Les algorithmes développés ici sont combinés à des listes d'inclusion, ce qui améliore significativement la détection d'isomères (séparés ou qui co-éluent en LC) et permet de localiser avec confiance le site modifié. Le fait de pouvoir identifier et localiser un évènement de phosphorylation précis est d'une importance essentielle quand vient le temps d'attribuer une fonction biologique à un site particulier.

Globalement, les objectifs que nous nous sommes fixés dans cette thèse ont été atteints et les défis futurs de ce projet sont décrits dans la section suivante.

5.3. Perspectives

Les travaux présentés dans cette thèse ont permis de recenser des milliers de sites de phosphorylations chez la drosophile dont la plupart n'avaient jamais été publiés auparavant. Contrairement à de nombreuses études parues durant la dernière décennie, les données présentées ici ne se contentent pas de servir uniquement de répertoire mais incluent également un aspect quantitatif évalué de manière statistique. Ainsi, ces renseignements peuvent servir de référence à d'autres études visant, par exemple, la fonction précise d'un site de phosphorylation dans un contexte similaire.

Nous avons identifié plusieurs événements de phosphorylation impliqués dans la voie de signalisation de l'insuline chez la drosophile ainsi que des cibles potentielles de CK2 suite au traitement par RNAi. Cependant, une investigation plus poussée de la fonction biologique associée à chacun des sites de phosphorylation modulés dans différentes conditions permettrait d'explorer plus en détail leur rôle physiologique. Des expériences de mutagenèse dirigée pourraient être conduites afin de muter les résidus d'intérêt et observer un phénotype associé. Il serait primordial de confirmer si les sites identifiés sont conservés et reproduire les observations obtenues chez la drosophile dans un modèle de culture cellulaire chez le mammifère. Ainsi, il serait envisageable de déterminer des cibles pharmacologiques applicables chez l'humain. Les travaux effectués permettraient d'appréhender les mécanismes physiologiques impliqués dans la voie de l'insuline et d'identifier de nouvelles cibles thérapeutiques pour le traitement du cancer ou de maladies métaboliques.

Il serait pertinent de s'intéresser plus en détails à l'analyse fonctionnelle de CK2 suite au traitement à l'insuline dans les cellules S2. Par exemple, il serait important de regarder l'effet de la localisation de CK2 sur son activité suite au traitement à l'insuline. Une étude phosphoprotéomique différentielle similaire pourrait être réalisée sur des fractions cytoplasmique et nucléaire afin d'avoir une idée plus précise des mécanismes d'action. Il serait également envisageable de confirmer les résultats des essais kinase *in vitro* sur les peptides synthétiques avec des essais similaires impliquant des protéines

recombinantes ayant une structure tridimensionnelle intacte. Il serait intéressant de se concentrer sur les protéines pouvant former un complexe avec CK2 afin d'identifier des changements de compositions ou de modifications post-traductionnelles suite au traitement à l'insuline. Cela pourrait nous donner des indices précieux sur l'activation/inactivation de protéines clés interagissant avec CK2 et régulant la voie de l'insuline.

Il serait intéressant d'étudier plus en détails le rôle d'InR dans la régulation d'enzymes responsables du stockage du glucose et de son utilisation chez la drosophile qui est un phénomène sûrement conservé. Ceci nous permettrait de découvrir de nouveaux composants impliqués dans la voie de l'insuline qui pourraient être de nouvelles cibles en vue du traitement du diabète.

D'une manière générale, il serait pertinent de valider les conclusions présentées dans cette thèse dans un contexte permettant d'appréhender les mécanismes intracellulaires régulés par l'insuline.

Deux algorithmes permettant la détection de phosphopeptides isomériques ont été présentés dans cette thèse. Il est essentiel de pouvoir identifier avec une haute confiance la localisation du site modifié afin d'associer une fonctionnalité précise à la phosphorylation. Cela devient primordial par exemple dans le cas où on veuille faire synthétiser un anticorps phosphospécifique très dispendieux afin de valider une cible d'intérêt. Nous pensons qu'il y aurait toujours place à l'amélioration afin de réduire le taux de faux positifs rapporté dans les analyses de balayage pour ensuite se concentrer sur les peptides d'intérêt ciblés dans la seconde analyse. Un perfectionnement de la programmation des scripts permettrait très certainement de lever certaines ambiguïtés et d'identifier davantage de cas de phosphoisomères avec une plus haute confiance. Il serait également envisageable d'analyser des phosphopeptides suite à un enrichissement grâce à la stratégie SIMAC (décrite en **section 1.7.4.4**) afin d'obtenir une plus grande diversité dans la population. Cela nous permettrait d'identifier davantage de peptides multiples phosphorylés ce qui multiplierait assurément le nombre d'isomères détectés. L'utilisation complémentaire de divers algorithmes permettant d'attribuer un score à la confiance de localisation du

phosphate sur le peptide serait bénéfique afin de compléter le A-score que nous avons utilisé. D'autres méthodes de séparation telles que la chromatographie HILIC ou le FAIMS seraient à tester. Il reste encore des optimisations techniques à réaliser afin de les utiliser en routine et les algorithmes devraient être adaptés pour la séparation de phosphopeptides isomériques par mobilité ionique. Les méthodes d'acquisition pourraient aussi être améliorées pour réaliser une étude plus exhaustive et couvrir une majorité de cas. Par exemple, le fait de diviser adéquatement les listes d'inclusion en plusieurs injections en se basant sur les temps de rétention des candidats permettrait d'identifier plus d'isomères. Ces algorithmes peuvent être utilisés par d'autres laboratoires dans le cas où un profil de phosphorylation complet est recherché pour une protéine particulière suite à une immunoprécipitation par exemple. Cela augmentera les chances d'identifier un maximum de sites présents sur la protéine grâce à la génération de listes d'inclusion. Suite à quelques ajustements, les scripts développés pourront également être applicables pour l'étude d'autres modifications post-traductionnelles.

Bibliographie

- [1] Bridon, G., Bonneil, E., Muratore-Schroeder, T., Caron-Lizotte, O., Thibault, P., Improvement of Phosphoproteome Analyses Using FAIMS and Decision Tree Fragmentation. Application to the Insulin Signaling Pathway in *Drosophila melanogaster* S2 Cells. *J Proteome Res* 2012, *11*, 927-940.
- [2] Courcelles, M., Bridon, G., Lemieux, S., Thibault, P., Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments. *J Proteome Res* 2012, *11*, 3753-3765.
- [3] Levene, P. A., Alsberg, C. L., The cleavage products of vitellin. *J. Biol. Chem.* 1906, *2*, 127 - 133.
- [4] Zhang, H., Zha, X., Tan, Y., Hornbeck, P. V., *et al.*, Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. *J Biol Chem* 2002, *277*, 39379-39387.
- [5] Hubbard, M. J., Cohen, P., On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem Sci* 1993, *18*, 172-177.
- [6] Dhanasekaran, N., Premkumar Reddy, E., Signaling by dual specificity kinases. *Oncogene* 1998, *17*, 1447-1455.
- [7] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., The protein kinase complement of the human genome. *Science* 2002, *298*, 1912-1934.
- [8] Higashiyama, S., Iwabuki, H., Morimoto, C., Hieda, M., *et al.*, Membrane-anchored growth factors, the epidermal growth factor family: beyond receptor ligands. *Cancer science* 2008, *99*, 214-220.
- [9] Schwartz, D., Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *23* 2005, *11*.
- [10] Shi, Y., Serine/threonine phosphatases: mechanism through structure. *Cell* 2009, *139*, 468-484.
- [11] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., *et al.*, The sequence of the human genome. *Science* 2001, *291*, 1304-1351.
- [12] Trost, M., Bridon, G., Desjardins, M., Thibault, P., Subcellular phosphoproteomics. *Mass Spectrom. Rev.* 2010, *29*, 962-990.
- [13] Lienhard, G. E., Non-functional phosphorylations? *Trends Biochem Sci* 2008, *33*, 351-352.
- [14] Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., *et al.*, PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* 2012, *40*, D261-270.
- [15] Boulikas, T., The phosphorylation connection to cancer (review). *Int J Oncol* 1995, *6*, 271-278.
- [16] Jorgensen, C., Sherman, A., Chen, G. I., Pasculescu, A., *et al.*, Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* 2009, *326*, 1502-1509.
- [17] Stokes, M. P., Rush, J., Macneill, J., Ren, J. M., *et al.*, Profiling of UV-induced ATM/ATR signaling pathways. *Proc Natl Acad Sci U S A* 2007, *104*, 19855-19860.

- [18] Bose, R., Molina, H., Patterson, A. S., Bitok, J. K., *et al.*, Phosphoproteomic analysis of Her2/neu signaling and inhibition. *Proc Natl Acad Sci U S A* 2006, 103, 9773-9778.
- [19] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, 127, 635-648.
- [20] Zanivan, S., Gnad, F., Wickstrom, S. A., Geiger, T., *et al.*, Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J Proteome Res* 2008, 7, 5314-5326.
- [21] Sharma, S. V., Bell, D. W., Settleman, J., Haber, D. A., Epidermal growth factor receptor mutations in lung cancer. *Nature reviews. Cancer* 2007, 7, 169-181.
- [22] Hynes, N. E., MacDonald, G., ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol* 2009, 21, 177-184.
- [23] Harsha, H. C., Jimeno, A., Molina, H., Mihalas, A. B., *et al.*, Activated epidermal growth factor receptor as a novel target in pancreatic cancer therapy. *J Proteome Res* 2008, 7, 4651-4658.
- [24] Lee, J. T., Lehmann, B. D., Terrian, D. M., Chappell, W. H., *et al.*, Targeting prostate cancer based on signal transduction and cell cycle pathways. *Cell cycle (Georgetown, Tex.)* 2008, 7, 1745-1762.
- [25] Corless, C. L., Fletcher, J. A., Heinrich, M. C., Biology of gastrointestinal stromal tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2004, 22, 3813-3825.
- [26] Ferrajoli, A., Faderl, S., Ravandi, F., Estrov, Z., The JAK-STAT pathway: a therapeutic target in hematological malignancies. *Current cancer drug targets* 2006, 6, 671-679.
- [27] Knight, Z. A., Lin, H., Shokat, K. M., Targeting the cancer kinome through polypharmacology. *Nature reviews. Cancer* 2010, 10, 130-137.
- [28] Pan, C., Olsen, J. V., Daub, H., Mann, M., Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol Cell Proteomics* 2009, 8, 2796-2808.
- [29] Xu, A. M., Huang, P. H., Receptor tyrosine kinase coactivation networks in cancer. *Cancer research* 2010, 70, 3857-3860.
- [30] Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., *et al.*, Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *The New England journal of medicine* 2004, 350, 2129-2139.
- [31] Meriggi, F., Zaniboni, A., Epidermal growth factor receptor tyrosine kinase inhibitors for elderly patients with advanced non-small cell lung cancer. *Current gerontology and geriatrics research* 2010, 348174.
- [32] Sloan, B., Scheinfeld, N. S., Pazopanib, a VEGF receptor tyrosine kinase inhibitor for cancer therapy. *Current opinion in investigational drugs (London, England : 2000)* 2008, 9, 1324-1335.
- [33] Stommel, J. M., Kimmelman, A. C., Ying, H., Nabioullin, R., *et al.*, Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* 2007, 318, 287-290.
- [34] Schneider, I., Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embryol Exp Morphol* 1972, 27, 353-365.

- [35] Rogers, S. L., Rogers, G. C., Culture of *Drosophila* S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nat Protoc* 2008, 3, 606-611.
- [36] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., *et al.*, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998, 391, 806-811.
- [37] Bernstein, E., Caudy, A. A., Hammond, S. M., Hannon, G. J., Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001, 409, 363-366.
- [38] Okamura, K., Ishizuka, A., Siomi, H., Siomi, M. C., Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 2004, 18, 1655-1666.
- [39] Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W., Sontheimer, E. J., A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* 2004, 117, 83-94.
- [40] Saleh, M. C., van Rij, R. P., Hekele, A., Gillis, A., *et al.*, The endocytic pathway mediates cell entry of dsRNA to induce RNAi silencing. *Nat Cell Biol* 2006, 8, 793-802.
- [41] Naito, Y., Yamada, T., Matsumiya, T., Ui-Tei, K., *et al.*, dsCheck: highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. *Nucleic Acids Res* 2005, 33, W589-591.
- [42] Saltiel, A. R., Kahn, C. R., Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* 2001, 414, 799-806.
- [43] Palanker, L., Tennessen, J. M., Lam, G., Thummel, C. S., *Drosophila* HNF4 regulates lipid mobilization and beta-oxidation. *Cell Metab* 2009, 9, 228-239.
- [44] Rulifson, E. J., Kim, S. K., Nusse, R., Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science* 2002, 296, 1118-1120.
- [45] Kim, S. K., Rulifson, E. J., Conserved mechanisms of glucose sensing and regulation by *Drosophila* corpora cardiaca cells. *Nature* 2004, 431, 316-320.
- [46] DiAngelo, J. R., Birnbaum, M. J., Regulation of fat cell mass by insulin in *Drosophila melanogaster*. *Mol Cell Biol* 2009, 29, 6341-6352.
- [47] Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., *et al.*, An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr Biol* 2001, 11, 213-221.
- [48] Clancy, D. J., Gems, D., Harshman, L. G., Oldham, S., *et al.*, Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* 2001, 292, 104-106.
- [49] Tatar, M., Kopelman, A., Epstein, D., Tu, M. P., *et al.*, A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* 2001, 292, 107-110.
- [50] Chen, C., Jack, J., Garofalo, R. S., The *Drosophila* insulin receptor is required for normal growth. *Endocrinology* 1996, 137, 846-856.
- [51] Fernandez-Almonacid, R., Rosen, O. M., Structure and ligand specificity of the *Drosophila melanogaster* insulin receptor. *Mol Cell Biol* 1987, 7, 2718-2727.

- [52] Ruan, Y., Chen, C., Cao, Y., Garofalo, R. S., The *Drosophila* insulin receptor contains a novel carboxyl-terminal extension likely to play an important role in signal transduction. *J Biol Chem* 1995, 270, 4236-4243.
- [53] Ogawa, W., Matozaki, T., Kasuga, M., Role of binding proteins to IRS-1 in insulin signalling. *Mol Cell Biochem* 1998, 182, 13-22.
- [54] Kwon, H. B., Kim, S. H., Kim, S. E., Jang, I. H., *et al.*, *Drosophila* extracellular signal-regulated kinase involves the insulin-mediated proliferation of Schneider cells. *J Biol Chem* 2002, 277, 14853-14858.
- [55] Skolnik, E. Y., Batzer, A., Li, N., Lee, C. H., *et al.*, The function of GRB2 in linking the insulin receptor to Ras signaling pathways. *Science* 1993, 260, 1953-1955.
- [56] Clemens, J. C., Worby, C. A., Simonson-Leff, N., Muda, M., *et al.*, Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc Natl Acad Sci U S A* 2000, 97, 6499-6503.
- [57] Zhang, W., Thompson, B. J., Hietakangas, V., Cohen, S. M., MAPK/ERK signaling regulates insulin sensitivity to control glucose metabolism in *Drosophila*. *PLoS Genet* 2011, 7, e1002429.
- [58] Tanoue, T., Nishida, E., Molecular recognitions in the MAP kinase cascades. *Cell Signal* 2003, 15, 455-462.
- [59] Ho, D. T., Bardwell, A. J., Grewal, S., Iverson, C., Bardwell, L., Interacting JNK-docking sites in MKK7 promote binding and activation of JNK mitogen-activated protein kinases. *J Biol Chem* 2006, 281, 13169-13179.
- [60] Bohni, R., Riesgo-Escovar, J., Oldham, S., Brogiolo, W., *et al.*, Autonomous control of cell and organ size by CHICO, a *Drosophila* homolog of vertebrate IRS1-4. *Cell* 1999, 97, 865-875.
- [61] Werz, C., Kohler, K., Hafen, E., Stocker, H., The *Drosophila* SH2B family adaptor Lnk acts in parallel to chico in the insulin signaling pathway. *PLoS Genet* 2009, 5, e1000596.
- [62] Taguchi, A., White, M. F., Insulin-like signaling, nutrient homeostasis, and life span. *Annu Rev Physiol* 2008, 70, 191-212.
- [63] Shepherd, P. R., Withers, D. J., Siddle, K., Phosphoinositide 3-kinase: the key switch mechanism in insulin signalling. *Biochem J* 1998, 333 (Pt 3), 471-490.
- [64] Scanga, S. E., Ruel, L., Binari, R. C., Snow, B., *et al.*, The conserved PI3'K/PTEN/Akt signaling pathway regulates both cell size and survival in *Drosophila*. *Oncogene* 2000, 19, 3971-3977.
- [65] Sarbassov, D. D., Guertin, D. A., Ali, S. M., Sabatini, D. M., Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* 2005, 307, 1098-1101.
- [66] Thomas, G., Hall, M. N., TOR signalling and control of cell growth. *Curr Opin Cell Biol* 1997, 9, 782-787.
- [67] Tapon, N., Ito, N., Dickson, B. J., Treisman, J. E., Hariharan, I. K., The *Drosophila* tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation. *Cell* 2001, 105, 345-355.

- [68] Puig, O., Marr, M. T., Ruhf, M. L., Tjian, R., Control of cell number by *Drosophila* FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev* 2003, 17, 2006-2020.
- [69] Asnaghi, L., Bruno, P., Priulla, M., Nicolin, A., mTOR: a protein kinase switching between life and death. *Pharmacological research : the official journal of the Italian Pharmacological Society* 2004, 50, 545-549.
- [70] Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M., Hemmings, B. A., Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* 1995, 378, 785-789.
- [71] McManus, E. J., Sakamoto, K., Armit, L. J., Ronaldson, L., *et al.*, Role that phosphorylation of GSK3 plays in insulin and Wnt signalling defined by knockin analysis. *EMBO J* 2005, 24, 1571-1583.
- [72] Doble, B. W., Woodgett, J. R., GSK-3: tricks of the trade for a multi-tasking kinase. *J Cell Sci* 2003, 116, 1175-1186.
- [73] Ferkey, D. M., Kimelman, D., Glycogen synthase kinase-3 beta mutagenesis identifies a common binding domain for GBP and Axin. *J Biol Chem* 2002, 277, 16147-16152.
- [74] Li, M., Wang, X., Meintzer, M. K., Laessig, T., *et al.*, Cyclic AMP promotes neuronal survival by phosphorylation of glycogen synthase kinase 3beta. *Mol Cell Biol* 2000, 20, 9356-9363.
- [75] Lizcano, J. M., Alessi, D. R., The insulin signalling pathway. *Curr Biol* 2002, 12, R236-238.
- [76] Loberg, R. D., Vesely, E., Brosius, F. C., 3rd, Enhanced glycogen synthase kinase-3beta activity mediates hypoxia-induced apoptosis of vascular smooth muscle cells and is prevented by glucose transport and metabolism. *J Biol Chem* 2002, 277, 41667-41673.
- [77] Leng, S., Zhang, W., Zheng, Y., Liberman, Z., *et al.*, Glycogen synthase kinase 3 beta mediates high glucose-induced ubiquitination and proteasome degradation of insulin receptor substrate 1. *The Journal of endocrinology* 2010, 206, 171-181.
- [78] Watson, R. T., Pessin, J. E., Intracellular organization of insulin signaling and GLUT4 translocation. *Recent progress in hormone research* 2001, 56, 175-193.
- [79] Jewell, J. L., Oh, E., Ramalingam, L., Kalwat, M. A., *et al.*, Munc18c phosphorylation by the insulin receptor links cell signaling directly to SNARE exocytosis. *J Cell Biol* 2011, 193, 185-199.
- [80] Schmelzle, K., Kane, S., Gridley, S., Lienhard, G. E., White, F. M., Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 2006, 55, 2171-2179.
- [81] Kruger, M., Kratchmarova, I., Blagoev, B., Tseng, Y. H., *et al.*, Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc Natl Acad Sci U S A* 2008, 105, 2451-2456.
- [82] Larance, M., Rowland, A. F., Hoehn, K. L., Humphreys, D. T., *et al.*, Global phosphoproteomics identifies a major role for AKT and 14-3-3 in regulating EDC3. *Mol Cell Proteomics* 2010, 9, 682-694.
- [83] Przybylski, C., Junger, M. A., Aubertin, J., Radvanyi, F., *et al.*, Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-Orbitrap instrument. *J Proteome Res* 2010, 9, 5118-5132.
- [84] Monetti, M., Nagaraj, N., Sharma, K., Mann, M., Large-scale phosphosite quantification in tissues by a spike-in SILAC method. *Nat Methods* 2011, 8, 655-658.

- [85] Filhol, O., Cochet, C., Protein kinase CK2 in health and disease: Cellular functions of protein kinase CK2: a dynamic affair. *Cell Mol Life Sci* 2009, *66*, 1830-1839.
- [86] Faust, M., Montenarh, M., Subcellular localization of protein kinase CK2. A key to its function? *Cell and tissue research* 2000, *301*, 329-340.
- [87] Donella-Deana, A., Cesaro, L., Sarno, S., Ruzzene, M., *et al.*, Tyrosine phosphorylation of protein kinase CK2 by Src-related tyrosine kinases correlates with increased catalytic activity. *Biochem J* 2003, *372*, 841-849.
- [88] Vilk, G., Weber, J. E., Turowec, J. P., Duncan, J. S., *et al.*, Protein kinase CK2 catalyzes tyrosine phosphorylation in mammalian cells. *Cell Signal* 2008, *20*, 1942-1951.
- [89] Tuazon, P. T., Traugh, J. A., Casein kinase I and II--multipotential serine protein kinases: structure, function, and regulation. *Adv Second Messenger Phosphoprotein Res* 1991, *23*, 123-164.
- [90] Hathaway, G. M., Traugh, J. A., Casein kinases--multipotential protein kinases. *Curr Top Cell Regul* 1982, *21*, 101-127.
- [91] Allende, J. E., Allende, C. C., Protein kinase CK2: an enzyme with multiple substrates and a puzzling regulation. *FASEB J* 1995, *9*, 313-323.
- [92] Guerra, B., Siemer, S., Boldyreff, B., Issinger, O. G., Protein kinase CK2: evidence for a protein kinase CK2beta subunit fraction, devoid of the catalytic CK2alpha subunit, in mouse brain and testicles. *FEBS Lett* 1999, *462*, 353-357.
- [93] Meggio, F., Marin, O., Pinna, L. A., Substrate specificity of protein kinase CK2. *Cell Mol Biol Res* 1994, *40*, 401-409.
- [94] Padmanabha, R., Chen-Wu, J. L., Hanna, D. E., Glover, C. V., Isolation, sequencing, and disruption of the yeast CKA2 gene: casein kinase II is essential for viability in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1990, *10*, 4089-4099.
- [95] Hanna, D. E., Rethinaswamy, A., Glover, C. V., Casein kinase II is required for cell cycle progression during G1 and G2/M in *Saccharomyces cerevisiae*. *J Biol Chem* 1995, *270*, 25905-25914.
- [96] Ahmed, K., Davis, A. T., Wang, H., Faust, R. A., *et al.*, Significance of protein kinase CK2 nuclear signaling in neoplasia. *Journal of cellular biochemistry. Supplement* 2000, *Suppl 35*, 130-135.
- [97] Trembley, J. H., Chen, Z., Unger, G., Slaton, J., *et al.*, Emergence of protein kinase CK2 as a key target in cancer therapy. *BioFactors (Oxford, England)* 2010, *36*, 187-195.
- [98] Guo, C., Yu, S., Davis, A. T., Wang, H., *et al.*, A potential role of nuclear matrix-associated protein kinase CK2 in protection against drug-induced apoptosis in cancer cells. *J Biol Chem* 2001, *276*, 5992-5999.
- [99] Dominguez, I., Sonenshein, G. E., Seldin, D. C., Protein kinase CK2 in health and disease: CK2 and its role in Wnt and NF-kappaB signaling: linking development and cancer. *Cell Mol Life Sci* 2009, *66*, 1850-1857.
- [100] Krippner-Heidenreich, A., Talanian, R. V., Sekul, R., Kraft, R., *et al.*, Targeting of the transcription factor Max during apoptosis: phosphorylation-regulated cleavage by caspase-5 at an unusual glutamic acid residue in position P1. *Biochem J* 2001, *358*, 705-715.

- [101] Litchfield, D. W., Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem J* 2003, *369*, 1-15.
- [102] Henriksson, M., Luscher, B., Proteins of the Myc network: essential regulators of cell growth and differentiation. *Advances in cancer research* 1996, *68*, 109-182.
- [103] Oelgeschlager, M., Krieg, J., Luscher-Firzlaff, J. M., Luscher, B., Casein kinase II phosphorylation site mutations in c-Myb affect DNA binding and transcriptional cooperativity with NF-M. *Mol Cell Biol* 1995, *15*, 5966-5974.
- [104] Lin, A., Frost, J., Deng, T., Smeal, T., *et al.*, Casein kinase II is a negative regulator of c-Jun DNA binding and AP-1 activity. *Cell* 1992, *70*, 777-789.
- [105] Meek, D. W., Simon, S., Kikkawa, U., Eckhart, W., The p53 tumour suppressor protein is phosphorylated at serine 389 by casein kinase II. *EMBO J* 1990, *9*, 3253-3260.
- [106] O'Brien, K. A., Lemke, S. J., Cocke, K. S., Rao, R. N., Beckmann, R. P., Casein kinase 2 binds to and phosphorylates BRCA1. *Biochem Biophys Res Commun* 1999, *260*, 658-664.
- [107] Bousset, K., Henriksson, M., Luscher-Firzlaff, J. M., Litchfield, D. W., Luscher, B., Identification of casein kinase II phosphorylation sites in Max: effects on DNA-binding kinetics of Max homo- and Myc/Max heterodimers. *Oncogene* 1993, *8*, 3211-3220.
- [108] Coqueret, O., Martin, N., Berube, G., Rabbat, M., *et al.*, DNA binding by cut homeodomain proteins is down-modulated by casein kinase II. *J Biol Chem* 1998, *273*, 2561-2566.
- [109] Lodie, T. A., Savedra, R., Jr., Golenbock, D. T., Van Beveren, C. P., *et al.*, Stimulation of macrophages by lipopolysaccharide alters the phosphorylation state, conformation, and function of PU.1 via activation of casein kinase II. *Journal of immunology (Baltimore, Md. : 1950)* 1997, *158*, 1848-1856.
- [110] Ford, H. L., Landesman-Bollag, E., Dacwag, C. S., Stukenberg, P. T., *et al.*, Cell cycle-regulated phosphorylation of the human SIX1 homeodomain protein. *J Biol Chem* 2000, *275*, 22245-22254.
- [111] Cardenas, M. E., Gasser, S. M., Regulation of topoisomerase II by phosphorylation: a role for casein kinase II. *J Cell Sci* 1993, *104 (Pt 2)*, 219-225.
- [112] Karandikar, U. C., Shaffer, J., Bishop, C. P., Bidwai, A. P., Drosophila CK2 phosphorylates Deadpan, a member of the HES family of basic-helix-loop-helix (bHLH) repressors. *Mol Cell Biochem* 2005, *274*, 133-139.
- [113] Kahali, B., Trott, R., Paroush, Z., Allada, R., *et al.*, Drosophila CK2 phosphorylates Hairy and regulates its activity in vivo. *Biochem Biophys Res Commun* 2008, *373*, 637-642.
- [114] Liu, Z. P., Galindo, R. L., Wasserman, S. A., A role for CKII phosphorylation of the cactus PEST domain in dorsoventral patterning of the Drosophila embryo. *Genes Dev* 1997, *11*, 3413-3422.
- [115] Bonet, C., Fernandez, I., Aran, X., Bernues, J., *et al.*, The GAGA protein of Drosophila is phosphorylated by CK2. *J Mol Biol* 2005, *351*, 562-572.
- [116] Nuthall, H. N., Joachim, K., Stifani, S., Phosphorylation of serine 239 of Groucho/TLE1 by protein kinase CK2 is important for inhibition of neuronal differentiation. *Mol Cell Biol* 2004, *24*, 8395-8407.

- [117] Siomi, M. C., Higashijima, K., Ishizuka, A., Siomi, H., Casein kinase II phosphorylates the fragile X mental retardation protein and modulates its biological properties. *Mol Cell Biol* 2002, 22, 8438-8447.
- [118] Li, M., Strand, D., Krehan, A., Pyerin, W., *et al.*, Casein kinase 2 binds and phosphorylates the nucleosome assembly protein-1 (NAP1) in *Drosophila melanogaster*. *J Mol Biol* 1999, 293, 1067-1084.
- [119] Zhao, T., Eissenberg, J. C., Phosphorylation of heterochromatin protein 1 by casein kinase II is required for efficient heterochromatin binding in *Drosophila*. *J Biol Chem* 1999, 274, 15095-15100.
- [120] Munstermann, U., Fritz, G., Seitz, G., Lu, Y. P., *et al.*, Casein kinase II is elevated in solid human tumours and rapidly proliferating non-neoplastic tissue. *Eur J Biochem* 1990, 189, 251-257.
- [121] Guerra, B., Issinger, O. G., Protein kinase CK2 in human diseases. *Current medicinal chemistry* 2008, 15, 1870-1886.
- [122] Cristiani, A., Costa, G., Cozza, G., Meggio, F., *et al.*, The role of the N-terminal domain in the regulation of the "constitutively active" conformation of protein kinase CK2alpha: insight from a molecular dynamics investigation. *ChemMedChem* 2011, 6, 1207-1216.
- [123] Olsen, B. B., Guerra, B., Niefind, K., Issinger, O. G., Structural basis of the constitutive activity of protein kinase CK2. *Methods Enzymol* 2010, 484, 515-529.
- [124] Olsten, M. E., Litchfield, D. W., Order or chaos? An evaluation of the regulation of protein kinase CK2. *Biochem Cell Biol* 2004, 82, 681-693.
- [125] Carroll, D., Marshak, D. R., Serum-stimulated cell growth causes oscillations in casein kinase II activity. *J Biol Chem* 1989, 264, 7345-7348.
- [126] Klarlund, J. K., Czech, M. P., Insulin-like growth factor I and insulin rapidly increase casein kinase II activity in BALB/c 3T3 fibroblasts. *J Biol Chem* 1988, 263, 15872-15875.
- [127] Sommercorn, J., Mulligan, J. A., Lozeman, F. J., Krebs, E. G., Activation of casein kinase II in response to insulin and to epidermal growth factor. *Proc Natl Acad Sci U S A* 1987, 84, 8834-8838.
- [128] Ackerman, P., Glover, C. V., Osheroff, N., Stimulation of casein kinase II by epidermal growth factor: relationship between the physiological activity of the kinase and the phosphorylation state of its beta subunit. *Proc Natl Acad Sci U S A* 1990, 87, 821-825.
- [129] Litchfield, D. W., Dobrowolska, G., Krebs, E. G., Regulation of casein kinase II by growth factors: a reevaluation. *Cell Mol Biol Res* 1994, 40, 373-381.
- [130] Meng, R., Gotz, C., Montenarh, M., The role of protein kinase CK2 in the regulation of the insulin production of pancreatic islets. *Biochem Biophys Res Commun* 2010, 401, 203-206.
- [131] Kalmykova, A. I., Dobritsa, A. A., Gvozdev, V. A., The Su(Ste) repeat in the Y chromosome and betaCK2tes gene encode predicted isoforms of regulatory beta-subunit of protein kinase CK2 in *Drosophila melanogaster*. *FEBS Lett* 1997, 416, 164-166.
- [132] Kalmykova, A. I., Dobritsa, A. A., Gvozdev, V. A., The beta CK2tes gene encodes the tissue-specific regulatory subunit of casein kinase 2 in *Drosophila melanogaster*. *Biochemistry (Mosc)* 1997, 62, 455-460.

- [133] Bidwai, A. P., Zhao, W., Glover, C. V., A gene located at 56F1-2 in *Drosophila melanogaster* encodes a novel metazoan beta-like subunit of casein kinase II. *Mol Cell Biol Res Commun* 1999, 1, 21-28.
- [134] Lin, W. J., Sheu, G. T., Traugh, J. A., Effects of autophosphorylation on casein kinase II activity: evidence from mutations in the beta subunit. *Biochemistry* 1994, 33, 6998-7004.
- [135] Jauch, E., Melzig, J., Brkulj, M., Raabe, T., In vivo functional analysis of *Drosophila* protein kinase casein kinase 2 (CK2) beta-subunit. *Gene* 2002, 298, 29-39.
- [136] Litchfield, D. W., Bosc, D. G., Slominski, E., The protein kinase from mitotic human cells that phosphorylates Ser-209 on the casein kinase II beta-subunit is p34cdc2. *Biochim Biophys Acta* 1995, 1269, 69-78.
- [137] Battistutta, R., De Moliner, E., Sarno, S., Zanotti, G., Pinna, L. A., Structural features underlying selective inhibition of protein kinase CK2 by ATP site-directed tetrabromo-2-benzotriazole. *Protein science : a publication of the Protein Society* 2001, 10, 2200-2206.
- [138] Chilin, A., Battistutta, R., Bortolato, A., Cozza, G., *et al.*, Coumarin as attractive casein kinase 2 (CK2) inhibitor scaffold: an integrate approach to elucidate the putative binding motif and explain structure-activity relationships. *Journal of medicinal chemistry* 2008, 51, 752-759.
- [139] Cozza, G., Mazzorana, M., Papinutto, E., Bain, J., *et al.*, Quinalizarin as a potent, selective and cell-permeable inhibitor of protein kinase CK2. *Biochem J* 2009, 421, 387-395.
- [140] Siddiqui-Jain, A., Drygin, D., Streiner, N., Chua, P., *et al.*, CX-4945, an orally bioavailable selective inhibitor of protein kinase CK2, inhibits prosurvival and angiogenic signaling and exhibits antitumor efficacy. *Cancer research* 2010, 70, 10288-10298.
- [141] Solares, A. M., Santana, A., Baladron, I., Valenzuela, C., *et al.*, Safety and preliminary efficacy data of a novel casein kinase 2 (CK2) peptide inhibitor administered intralesionally at four dose levels in patients with cervical malignancies. *BMC cancer* 2009, 9, 146.
- [142] Trost, M., English, L., Lemieux, S., Courcelles, M., *et al.*, The phagosomal proteome in interferon-g-activated macrophages. *Immunity* 2009, 30, 143-154.
- [143] Thingholm, T. E., Larsen, M. R., Ingrell, C. R., Kassem, M., Jensen, O. N., TiO(2)-based phosphoproteomic analysis of the plasma membrane and the effects of phosphatase inhibitor treatment. *J Proteome Res* 2008, 7, 3304-3313.
- [144] Reinders, J., Sickmann, A., State-of-the-art in phosphoproteomics. *Proteomics* 2005, 5, 4052-4061.
- [145] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20, 3551-3567.
- [146] Swaney, D. L., Wenger, C. D., Coon, J. J., Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* 2010, 9, 1323-1329.
- [147] Candiano, G., Bruschi, M., Musante, L., Santucci, L., *et al.*, Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* 2004, 25, 1327-1333.

- [148] Gharib, M., Marcantonio, M., Lehmann, S. G., Courcelles, M., *et al.*, Artfactual sulfation of silver-stained proteins: implications for the assignment of phosphorylation and sulfation sites. *Mol Cell Proteomics* 2009, 8, 506-518.
- [149] Smejkal, G. B., Robinson, M. H., Lazarev, A., Comparison of fluorescent stains: relative photostability and differential staining of proteins in two-dimensional gels. *Electrophoresis* 2004, 25, 2511-2519.
- [150] Gauci, V. J., Wright, E. P., Coorsen, J. R., Quantitative proteomics: assessing the spectrum of in-gel protein detection methods. *Journal of chemical biology* 2011, 4, 3-29.
- [151] Steinberg, T. H., Protein gel staining methods: an introduction and overview. *Methods Enzymol* 2009, 463, 541-563.
- [152] Brumbaugh, K., Johnson, W., Liao, W. C., Lin, M. S., *et al.*, Overview of the generation, validation, and application of phosphosite-specific antibodies. *Methods Mol Biol* 2011, 717, 3-43.
- [153] Jacob, A. M., Turck, C. W., Detection of post-translational modifications by fluorescent staining of two-dimensional gels. *Methods Mol Biol* 2008, 446, 21-32.
- [154] Su, H. C., Hutchison, C. A., 3rd, Giddings, M. C., Mapping phosphoproteins in *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. *BMC microbiology* 2007, 7, 63.
- [155] Eymann, C., Becher, D., Bernhardt, J., Gronau, K., *et al.*, Dynamics of protein phosphorylation on Ser/Thr/Tyr in *Bacillus subtilis*. *Proteomics* 2007, 7, 3509-3526.
- [156] Geahlen, R. L., Anostario, M., Jr., Low, P. S., Harrison, M. L., Detection of protein kinase activity in sodium dodecyl sulfate-polyacrylamide gels. *Anal Biochem* 1986, 153, 151-158.
- [157] Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., *et al.*, A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* 2002, 1, 517-527.
- [158] Collins, M. O., Yu, L., Choudhary, J. S., Analysis of protein phosphorylation on a proteome-scale. *Proteomics* 2007, 7, 2751-2768.
- [159] Steen, H., Kuster, B., Fernandez, M., Pandey, A., Mann, M., Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* 2002, 277, 1031-1039.
- [160] Rush, J., Moritz, A., Lee, K. A., Guo, A., *et al.*, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 2005, 23, 94-101.
- [161] Villen, J., Beausoleil, S. A., Gerber, S. A., Gygi, S. P., Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* 2007, 104, 1488-1493.
- [162] Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R., 3rd, *et al.*, ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 2007, 316, 1160-1166.
- [163] Ndassa, Y. M., Orsi, C., Marto, J. A., Chen, S., Ross, M. M., Improved immobilized metal affinity chromatography for large-scale phosphoproteomics applications. *J Proteome Res* 2006, 5, 2789-2799.

- [164] Villen, J., Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* 2008, 3, 1630-1638.
- [165] Guerrero, I. C., Predic-Atkinson, J., Kleiner, O., Soskic, V., Godovac-Zimmermann, J., Enrichment of phosphoproteins for proteomic analysis using immobilized Fe(III)-affinity adsorption chromatography. *J Proteome Res* 2005, 4, 1545-1553.
- [166] Andersson, L., Porath, J., Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography. *Anal Biochem* 1986, 154, 250-254.
- [167] Li, S., Dass, C., Iron(III)-immobilized metal ion affinity chromatography and mass spectrometry for the purification and characterization of synthetic phosphopeptides. *Anal Biochem* 1999, 270, 9-14.
- [168] Posewitz, M. C., Tempst, P., Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* 1999, 71, 2883-2892.
- [169] Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J., *et al.*, Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 2002, 20, 301-305.
- [170] Jensen, S. S., Larsen, M. R., Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Commun Mass Spectrom* 2007, 21, 3635-3645.
- [171] Thingholm, T. E., Jensen, O. N., Robinson, P. J., Larsen, M. R., SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics* 2008, 7, 661-671.
- [172] Sano, A., Nakamura, H., Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. *Anal Sci* 2004, 20, 565-566.
- [173] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 2005, 4, 873-886.
- [174] Sugiyama, N., Masuda, T., Shinoda, K., Nakamura, A., *et al.*, Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol Cell Proteomics* 2007, 6, 1103-1109.
- [175] Mohammed, S., Kraiczek, K., Pinkse, M. W., Lemeer, S., *et al.*, Chip-Based Enrichment and NanoLC-MS/MS Analysis of Phosphopeptides from Whole Lysates. *J Proteome Res* 2008, 7, 1565-1571.
- [176] Raijmakers, R., Kraiczek, K., de Jong, A. P., Mohammed, S., Heck, A. J., Exploring the human leukocyte phosphoproteome using a microfluidic reversed-phase-TiO₂-reversed-phase high-performance liquid chromatography phosphochip coupled to a quadrupole time-of-flight mass spectrometer. *Anal Chem* 2010, 82, 824-832.
- [177] Chang, C. K., Wu, C. C., Wang, Y. S., Chang, H. C., Selective extraction and enrichment of multiphosphorylated peptides using polyarginine-coated diamond nanoparticles. *Anal Chem* 2008, 80, 3791-3797.
- [178] Shiau, K. J., Hung, S. U., Lee, H. W., Wu, C. C., Nanodiamond-based two-step sampling of multiply and singly phosphorylated peptides for MALDI-TOF mass spectrometry analysis. *Analyst* 2011, 136, 1922-1927.

- [179] Iliuk, A. B., Martin, V. A., Alicie, B. M., Geahlen, R. L., Tao, W. A., In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanoparticles. *Mol Cell Proteomics* 2010, 9, 2162-2172.
- [180] Iliuk, A., Jayasundera, K., Schluttenhofer, R., Tao, W. A., Functionalized soluble nanoparticles for phosphoproteome analysis. *Methods Mol Biol* 2011, 790, 277-285.
- [181] Zhou, H., Tian, R., Ye, M., Xu, S., *et al.*, Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. *Electrophoresis* 2007, 28, 2201-2215.
- [182] Wolschin, F., Wienkoop, S., Weckwerth, W., Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics* 2005, 5, 4389-4397.
- [183] Ficarro, S. B., Parikh, J. R., Blank, N. C., Marto, J. A., Niobium(V) oxide (Nb₂O₅): application to phosphoproteomics. *Anal Chem* 2008, 80, 4606-4613.
- [184] Liang, X., Fonnum, G., Hajivandi, M., Stene, T., *et al.*, Quantitative comparison of IMAC and TiO₂ surfaces used in the study of regulated, dynamic protein phosphorylation. *J Am Soc Mass Spectrom* 2007, 18, 1932-1944.
- [185] Wei, J., Zhang, Y., Wang, J., Tan, F., *et al.*, Highly efficient enrichment of phosphopeptides by magnetic nanoparticles coated with zirconium phosphonate for phosphoproteome analysis. *Rapid Commun Mass Spectrom* 2008, 22, 1069-1080.
- [186] Qiao, L., Roussel, C., Wan, J., Yang, P., *et al.*, Specific on-plate enrichment of phosphorylated peptides for direct MALDI-TOF MS analysis. *J Proteome Res* 2007, 6, 4763-4769.
- [187] Torta, F., Fusi, M., Casari, C. S., Bassi, A. L., Bachi, A., Nanostructured TiO thin films for phosphoproteomics studies with MALDI mass spectrometry. *Methods Mol Biol* 2011, 790, 173-181.
- [188] Reynolds, E. C., Riley, P. F., Adamson, N. J., A selective precipitation purification procedure for multiple phosphoserine-containing peptides and methods for their identification. *Anal Biochem* 1994, 217, 277-284.
- [189] Zhang, X., Ye, J., Jensen, O. N., Roepstorff, P., Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment. *Mol Cell Proteomics* 2007, 6, 2032-2042.
- [190] Oda, Y., Nagasu, T., Chait, B. T., Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001, 19, 379-382.
- [191] McLachlin, D. T., Chait, B. T., Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem* 2003, 75, 6826-6836.
- [192] Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., *et al.*, Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* 2003, 21, 1047-1054.
- [193] Zhou, H., Watts, J. D., Aebersold, R., A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* 2001, 19, 375-378.

- [194] Tao, W. A., Wollscheid, B., O'Brien, R., Eng, J. K., *et al.*, Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat Methods* 2005, 2, 591-598.
- [195] Bodenmiller, B., Mueller, L. N., Pedrioli, P. G., Pflieger, D., *et al.*, An integrated chemical, mass spectrometric and computational strategy for (quantitative) phosphoproteomics: application to *Drosophila melanogaster* Kc167 cells. *Molecular bioSystems* 2007, 3, 275-286.
- [196] Warthaka, M., Karwowska-Desaulniers, P., Pflum, M. K., Phosphopeptide modification and enrichment by oxidation-reduction condensation. *ACS chemical biology* 2006, 1, 697-701.
- [197] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004, 101, 12130-12135.
- [198] Trinidad, J. C., Specht, C. G., Thalhammer, A., Schoepfer, R., Burlingame, A. L., Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol Cell Proteomics* 2006, 5, 914-922.
- [199] Courcelles, M., Lemieux, S., Voisin, L., Meloche, S., Thibault, P., ProteoConnections: a bioinformatics platform to facilitate proteome and phosphoproteome analyses. *Proteomics* 2011, 11, 2654-2671.
- [200] Zarei, M., Sprenger, A., Metzger, F., Gretzmeier, C., Dengjel, J., Comparison of ERLIC-TiO₂, HILIC-TiO₂, and SCX-TiO₂ for global phosphoproteomics approaches. *J Proteome Res* 2011, 10, 3474-3483.
- [201] Mann, M., Ong, S. E., Gronborg, M., Steen, H., *et al.*, Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends in biotechnology* 2002, 20, 261-268.
- [202] Dzidic, I., Carroll, D. I., Stillwell, R. N., Horning, E. C., Atmospheric pressure ionization (API) mass spectrometry: formation of phenoxide ions from chlorinated aromatic compounds. *Anal Chem* 1975, 47, 1308-1312.
- [203] Morris, H. R., Panico, M., Barber, M., Bordoli, R. S., *et al.*, Fast atom bombardment: a new mass spectrometric method for peptide sequence analysis. *Biochem Biophys Res Commun* 1981, 101, 623-631.
- [204] Takats, Z., Wiseman, J. M., Gologan, B., Cooks, R. G., Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* 2004, 306, 471-473.
- [205] Magee, C. W., Harrington, W. L., Honig, R. E., Secondary ion quadrupole mass spectrometer for depth profiling--design and performance evaluation. *The Review of scientific instruments* 1978, 49, 477.
- [206] Macfarlane, R. D., Torgerson, D. F., Californium-252 plasma desorption mass spectroscopy. *Science* 1976, 191, 920-925.
- [207] Whitehouse, C. M., Dreyer, R. N., Yamashita, M., Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* 1985, 57, 675-679.
- [208] Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* 2000, 35, 804-817.
- [209] Bothner, B., Siuzdak, G., Electrospray ionization of a whole virus: analyzing mass, structure, and viability. *Chembiochem : a European journal of chemical biology* 2004, 5, 258-260.

- [210] Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000, 72, 1156-1162.
- [211] Michalski, A., Damoc, E., Lange, O., Denisov, E., *et al.*, Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Mol Cell Proteomics* 2012, 11, O111 013698.
- [212] Gorshkov, M. P., USSR Inventors Certificate No. 966583. 1982.
- [213] Purves, R. W., Guevremont, R., Day, S., Pipich, C. W., Matysjaszczyk, M. S., Mass spectrometric characterization of a high-field asymmetric waveform ion mobility spectrometer. *Rev. Sci. Instrum.* 1998, 69, 4094-4105.
- [214] Ibrahim, Y. M., Shvartsburg, A. A., Smith, R. D., Belov, M. E., Ultrasensitive identification of localization variants of modified peptides using ion mobility spectrometry. *Anal Chem* 2011, 83, 5617-5623.
- [215] Shvartsburg, A. A., Clemmer, D. E., Smith, R. D., Isotopic effect on ion mobility and separation of isotopomers by high-field ion mobility spectrometry. *Anal Chem* 2010, 82, 8047-8051.
- [216] Shvartsburg, A. A., Creese, A. J., Smith, R. D., Cooper, H. J., Separation of peptide isomers with variant modified sites by high-resolution differential ion mobility spectrometry. *Anal Chem* 2010, 82, 8327-8334.
- [217] Shvartsburg, A. A., Creese, A. J., Smith, R. D., Cooper, H. J., Separation of a set of peptide sequence isomers using differential ion mobility spectrometry. *Anal Chem* 2011, 83, 6918-6923.
- [218] Shvartsburg, A. A., Singer, D., Smith, R. D., Hoffmann, R., Ion mobility separation of isomeric phosphopeptides from a protein with variant modification of adjacent residues. *Anal Chem* 2011, 83, 5078-5085.
- [219] Mabrouki, R., Kelly, R. T., Prior, D. C., Shvartsburg, A. A., *et al.*, Improving FAIMS sensitivity using a planar geometry with slit interfaces. *J Am Soc Mass Spectrom* 2009, 20, 1768-1774.
- [220] Shvartsburg, A. A., Li, F., Tang, K., Smith, R. D., High-resolution field asymmetric waveform ion mobility spectrometry using new planar geometry analyzers. *Anal Chem* 2006, 78, 3706-3714.
- [221] Shvartsburg, A. A., Prior, D. C., Tang, K., Smith, R. D., High-resolution differential ion mobility separations using planar analyzers at elevated dispersion fields. *Anal Chem* 2010, 82, 7649-7655.
- [222] Shvartsburg, A. A., Smith, R. D., Scaling of the resolving power and sensitivity for planar FAIMS and mobility-based discrimination in flow- and field-driven analyzers. *J Am Soc Mass Spectrom* 2007, 18, 1672-1681.
- [223] Shvartsburg, A. A., Mashkevich, S. V., Smith, R. D., Feasibility of higher-order differential ion mobility separations using new asymmetric waveforms. *J Phys Chem A* 2006, 110, 2663-2673.
- [224] Shvartsburg, A. A., Smith, R. D., Optimum waveforms for differential ion mobility spectrometry (FAIMS). *J Am Soc Mass Spectrom* 2008, 19, 1286-1295.
- [225] Prieto, M., Tsai, C. W., Boumsellek, S., Ferran, R., *et al.*, Comparison of rectangular and bisinusoidal waveforms in a miniature planar high-field asymmetric waveform ion mobility spectrometer. *Anal Chem* 2011, 83, 9237-9243.

- [226] Barnett, D. A., Ells, B., Guevremont, R., Purves, R. W., Viehland, L. A., Evaluation of carrier gases for use in high-field asymmetric waveform ion mobility spectrometry. *J Am Soc Mass Spectrom* 2000, *11*, 1125-1133.
- [227] Shvartsburg, A. A., Danielson, W. F., Smith, R. D., High-resolution differential ion mobility separations using helium-rich gases. *Anal Chem* 2010, *82*, 2456-2462.
- [228] Shvartsburg, A. A., Smith, R. D., Accelerated high-resolution differential ion mobility separations using hydrogen. *Anal Chem* 2011, *83*, 9159-9166.
- [229] Barnett, D. A., Ouellette, R. J., Elimination of the helium requirement in high-field asymmetric waveform ion mobility spectrometry (FAIMS): beneficial effects of decreasing the analyzer gap width on peptide analysis. *Rapid Commun Mass Spectrom* 2011, *25*, 1959-1971.
- [230] Barnett, D. A., Ells, B., Guevremont, R., Purves, R. W., Application of ESI-FAIMS-MS to the analysis of tryptic peptides. *J Am Soc Mass Spectrom* 2002, *13*, 1282-1291.
- [231] Barnett, D. A., Ding, L., Ells, B., Purves, R. W., Guevremont, R., Tandem mass spectra of tryptic peptides at signal-to-background ratios approaching unity using electrospray ionization high-field asymmetric waveform ion mobility spectrometry/hybrid quadrupole time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2002, *16*, 676-680.
- [232] Swearingen, K. E., Hoopmann, M. R., Johnson, R. S., Saleem, R. A., *et al.*, Nanospray FAIMS fractionation provides significant increases in proteome coverage of unfractionated complex protein digests. *Mol Cell Proteomics* 2011.
- [233] Saba, J., Bonneil, E., Pomiès, C., Eng, K., Thibault, P., Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear ion trap/Orbitrap mass spectrometer. *J. Proteome Res.* 2009, *2009*, 3355-3366.
- [234] Venne, K., Bonneil, E., Eng, K., Thibault, P., Improvement in peptide detection for proteomics analyses using NanoLC-MS and high-field asymmetry waveform ion mobility mass spectrometry *Anal. Chem.* 2005, *77*, 2176-2186.
- [235] Brown, L. J., Toutoungi, D. E., Devenport, N. A., Reynolds, J. C., *et al.*, Miniaturized ultra high field asymmetric waveform ion mobility spectrometry combined with mass spectrometry for peptide analysis. *Anal Chem* 2010, *82*, 9827-9834.
- [236] Li, H., Giles, K., Bendiak, B. K., Kaplan, K. A., *et al.*, Resolving Structural Isomers of Monosaccharide Methyl Glycosides Using Drift Tube and Traveling Wave Ion Mobility Mass Spectrometry. *Anal Chem* 2012.
- [237] Edman, P., A method for the determination of amino acid sequence in peptides. *Archives of biochemistry* 1949, *22*, 475.
- [238] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* 1984, *11*, 601.
- [239] Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomedical & environmental mass spectrometry* 1988, *16*, 99-111.
- [240] Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology* 2004, *5*, 699-711.

- [241] Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N., Kirschner, M. W., Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics* 2006, 5, 172-181.
- [242] Ulintz, P. J., Yocum, A. K., Bodenmiller, B., Aebersold, R., *et al.*, Comparison of MS(2)-only, MSA, and MS(2)/MS(3) methodologies for phosphopeptide identification. *J Proteome Res* 2009, 8, 887-899.
- [243] Marcantonio, M., Trost, M., Courcelles, M., Desjardins, M., 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, 645-660.
- [244] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004, 101, 9528-9533.
- [245] Wiesner, J., Premisler, T., Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* 2008, 8, 4466-4483.
- [246] Mikesch, L. M., Ueberheide, B., Chi, A., Coon, J. J., *et al.*, The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* 2006, 1764, 1811-1822.
- [247] Swaney, D. L., McAlister, G. C., Coon, J. J., Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* 2008, 5, 959-964.
- [248] Olsen, J. V., Macek, B., Lange, O., Makarov, A., *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 2007, 4, 709-712.
- [249] Nagaraj, N., D'Souza, R. C., Cox, J., Olsen, J. V., Mann, M., Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res* 2010, 9, 6786-6794.
- [250] Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* 2003, 100, 6940-6945.
- [251] Rivers, J., Simpson, D. M., Robertson, D. H., Gaskell, S. J., Beynon, R. J., Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol Cell Proteomics* 2007, 6, 1416-1427.
- [252] Brun, V., Masselon, C., Garin, J., Dupuis, A., Isotope dilution strategies for absolute quantitative proteomics. *Journal of proteomics* 2009, 72, 740-749.
- [253] Brun, V., Dupuis, A., Adrait, A., Marcellin, M., *et al.*, Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics* 2007, 6, 2139-2149.
- [254] Chelius, D., Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* 2002, 1, 317-323.
- [255] Chalkley, R. J., Hansen, K. C., Baldwin, M. A., Bioinformatic methods to exploit mass spectrometric data for proteomic applications. *Methods Enzymol* 2005, 402, 289-312.

- [256] Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., *et al.*, Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 2005, 4, 1487-1502.
- [257] Fermin, D., Basrur, V., Yocum, A. K., Nesvizhskii, A. I., Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* 2011, 11, 1340-1345.
- [258] Braisted, J. C., Kuntumalla, S., Vogel, C., Marcotte, E. M., *et al.*, The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* 2008, 9, 529.
- [259] Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., *et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002, 1, 376-386.
- [260] Ong, S. E., Kratchmarova, I., Mann, M., Properties of ¹³C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* 2003, 2, 173-181.
- [261] Geiger, T., Wisniewski, J. R., Cox, J., Zanivan, S., *et al.*, Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat Protoc* 2011, 6, 147-157.
- [262] Yao, X., Freas, A., Ramirez, J., Demirev, P. A., Fenselau, C., Proteolytic ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 2001, 73, 2836-2842.
- [263] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., *et al.*, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, 17, 994-999.
- [264] Shadforth, I. P., Dunkley, T. P., Lilley, K. S., Bessant, C., i-Tracker: for quantitative proteomics using iTRAQ. *BMC genomics* 2005, 6, 145.
- [265] Thompson, A., Schafer, J., Kuhn, K., Kienle, S., *et al.*, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 2003, 75, 1895-1904.
- [266] Wu, R., Dephoure, N., Haas, W., Huttlin, E. L., *et al.*, Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. *Mol Cell Proteomics* 2011, 10, M111 009654.
- [267] Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003, 2, 43-50.
- [268] Nesvizhskii, A. I., Aebersold, R., Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug discovery today* 2004, 9, 173-181.
- [269] Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., Schieltz, D., Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 1995, 67, 1426-1436.
- [270] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics (Oxford, England)* 2004, 20, 1466-1467.

- [271] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008, 26, 1367-1372.
- [272] Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., Gygi, S. P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 2006, 24, 1285-1292.
- [273] Wan, Y., Cripps, D., Thomas, S., Campbell, P., *et al.*, PhosphoScan: a probability-based method for phosphorylation site prediction using MS2/MS3 pair information. *J Proteome Res* 2008, 7, 2803-2811.
- [274] Olsen, J. V., Mann, M., Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A* 2004, 101, 13417-13422.
- [275] Ruttenberg, B. E., Pisitkun, T., Knepper, M. A., Hoffert, J. D., PhosphoScore: an open-source phosphorylation site assignment tool for MSn data. *J Proteome Res* 2008, 7, 3054-3059.
- [276] MacLean, D., Burrell, M. A., Studholme, D. J., Jones, A. M., PhosCalc: a tool for evaluating the sites of peptide phosphorylation from mass spectrometer data. *BMC Res Notes* 2008, 1, 30.
- [277] Bailey, C. M., Sweet, S. M., Cunningham, D. L., Zeller, M., *et al.*, SLoMo: automated site localization of modifications from ETD/ECD mass spectra. *J Proteome Res* 2009, 8, 1965-1971.
- [278] Swaney, D. L., Wenger, C. D., Thomson, J. A., Coon, J. J., Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 2009, 106, 995-1000.
- [279] Taus, T., Kocher, T., Pichler, P., Paschke, C., *et al.*, Universal and Confident Phosphorylation Site Localization Using phosphoRS. *J Proteome Res* 2011, 10, 5354-5362.
- [280] Lemeer, S., Kunold, E., Klaeger, S., Raabe, M., *et al.*, Phosphorylation site localization in peptides by MALDI MS/MS and the Mascot Delta Score. *Analytical and bioanalytical chemistry* 2012, 402, 249-260.
- [281] Baker, P. R., Trinidad, J. C., Chalkley, R. J., Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics* 2011, 10, M111 008078.
- [282] Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., Zhang, B., PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 2004, 4, 1551-1561.
- [283] Gnad, F., Ren, S., Cox, J., Olsen, J. V., *et al.*, PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome biology* 2007, 8, R250.
- [284] Bodenmiller, B., Campbell, D., Gerrits, B., Lam, H., *et al.*, PhosphoPep—a database of protein phosphorylation sites in model organisms. *Nat Biotechnol* 2008, 26, 1339-1340.
- [285] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., *et al.*, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003, 31, 365-370.
- [286] Diella, F., Gould, C. M., Chica, C., Via, A., Gibson, T. J., Phospho.ELM: a database of phosphorylation sites—update 2008. *Nucleic Acids Res* 2008, 36, D240-244.

- [287] Ritz, A., Shakhnarovich, G., Salomon, A. R., Raphael, B. J., Discovery of phosphorylation motif mixtures in phosphoproteomics data. *Bioinformatics (Oxford, England)* 2009, 25, 14-21.
- [288] Bairoch, A., The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acids Res* 1993, 21, 3097-3103.
- [289] Blom, N., Gammeltoft, S., Brunak, S., Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999, 294, 1351-1362.
- [290] Huang, H. D., Lee, T. Y., Tzeng, S. W., Horng, J. T., KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res* 2005, 33, W226-229.
- [291] Xue, Y., Li, A., Wang, L., Feng, H., Yao, X., PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics* 2006, 7, 163.
- [292] Kim, J. H., Lee, J., Oh, B., Kimm, K., Koh, I., Prediction of phosphorylation sites using SVMs. *Bioinformatics (Oxford, England)* 2004, 20, 3179-3184.
- [293] Obenaus, J. C., Cantley, L. C., Yaffe, M. B., Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* 2003, 31, 3635-3641.
- [294] Linding, R., Jensen, L. J., Pasculescu, A., Olhovskiy, M., *et al.*, NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acid Res.* 2008, 36, D695-699.
- [295] Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., *et al.*, InterPro: the integrative protein signature database. *Nucleic Acids Res* 2009, 37, D211-215.
- [296] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., *et al.*, The Protein Data Bank. *Nucleic Acids Res* 2000, 28, 235-242.
- [297] Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., *et al.*, The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011, 39, D561-568.
- [298] Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., Ideker, T., Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics (Oxford, England)* 2011, 27, 431-432.
- [299] Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., *et al.*, PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* 2010, 38, D204-210.
- [300] Carbon, S., Ireland, A., Mungall, C. J., Shu, S., *et al.*, AmiGO: online access to ontology and annotation data. *Bioinformatics (Oxford, England)* 2009, 25, 288-289.
- [301] Ballardini, R., Benevento, M., Arrigoni, G., Pattini, L., Roda, A., MassUntangler: a novel alignment tool for label-free liquid chromatography-mass spectrometry proteomic data. *J Chromatogr A* 2011, 1218, 8859-8868.
- [302] Colaert, N., Gevaert, K., Martens, L., RIBAR and xRIBAR: Methods for reproducible relative MS/MS-based label-free protein quantification. *J Proteome Res* 2011, 10, 3183-3189.
- [303] Park, S. K., Venable, J. D., Xu, T., Yates, J. R., 3rd, A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods* 2008, 5, 319-322.

- [304] Bridges, S. M., Magee, G. B., Wang, N., Williams, W. P., *et al.*, ProtQuant: a tool for the label-free quantification of MudPIT proteomics data. *BMC Bioinformatics* 2007, 8 Suppl 7, S24.
- [305] Kleinnijenhuis, A. J., Kjeldsen, F., Kallipolitis, B., Haselmann, K. F., Jensen, O. N., Analysis of histidine phosphorylation using tandem MS and ion-electron reactions. *Anal Chem* 2007, 79, 7450-7456.
- [306] Brognard, J., Hunter, T., Protein kinase signaling networks in cancer. *Current opinion in genetics & development* 2011, 21, 4-11.
- [307] Moser, K., White, F. M., Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. *J. Proteome Res.* 2006, 5, 98-104.
- [308] Ndassa, Y. M., Orsi, C., Marto, J. A., Chen, S., Ross, M. M., Improved immobilized metal affinity chromatography for large-scale phosphoproteomics applications. *J. Proteome Res.* 2006, 5, 2789-2799.
- [309] Tsai, C. F., Wang, Y. T., Chen, Y. R., Lai, C. Y., *et al.*, Immobilized metal affinity chromatography revisited: pH/acid control toward high selectivity in phosphoproteomics. *J. Proteome Res.* 2008, 7, 4058-4069.
- [310] Ye, J., Zhang, X., Young, C., Zhao, X., *et al.*, Optimized IMAC-IMAC protocol for phosphopeptide recovery from complex biological samples. *J. Proteome Res.* 2010, 9, 3561-3573.
- [311] Hilger, M., Bonaldi, T., Gnad, F., Mann, M., System-wide analysis of a phosphatase knock down by quantitative proteomics and phosphoproteomics. *Mol. Cell. Proteomics* 2009, 8, 1908-1920.
- [312] Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., *et al.*, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, 127, 635-648.
- [313] Lemeer, S., Heck, A. J., The phosphoproteomics data explosion. *Current opinion in chemical biology* 2009, 13, 414-420.
- [314] Højlund, K., Bowen, B. P., Hwang, H., Flynn, C. R., *et al.*, In vivo phosphoproteome of human skeletal muscle revealed by phosphopeptide enrichment and HPLC-ESI-MS/MS. *J. Proteome Res.* 2009, 8, 4954-4965.
- [315] Nie, S., Dai, J., Ning, Z. B., Cao, X. J., *et al.*, Comprehensive Profiling of Phosphopeptides Based on Anion Exchange Followed by Flow-Through Enrichment with Titanium Dioxide (AFET) *J. Proteome Res.* 2010, 9, 4585-4594.
- [316] Swaney, D. L., McAlister, G. C., Coon, J. J., Decision tree driven tandem mass spectrometry for shotgun proteomics. *Nat. Methods* 2008, 5, 959-964.
- [317] Canterbury, J. D., Yi, X., Hoopmann, M. R., McCoss, M. J., Assessing the dynamic range and peak capacity of nanoflow LC-FAIMS-MS on an ion trap mass spectrometer for proteomics. *Anal. Chem.* 2008, 80, 6888-6897.
- [318] Merenbloom, S. I., Koeniger, S. L., Valentine, S. J., Plasencia, M. D., Clemmer, D. E., IMS-IMS and IMS-IMS-IMS/MS for separating peptide and protein fragment ions. *Anal Chem* 2006, 78, 2802-2809.
- [319] Valentine, S. J., Ewing, M. A., Dilger, J. M., Glover, M. S., *et al.*, Using ion mobility data to improve peptide identification: intrinsic amino acid size parameters. *J Proteome Res* 2011, 10, 2318-2329.

- [320] Tang, K., Li, F., Shvartsburg, A. A., Strittmatter, E. F., Smith, R. D., Two-dimensional gas-phase separations coupled to mass spectrometry for analysis of complex mixtures. *Anal Chem* 2005, 77, 6381-6388.
- [321] Xuan, Y., Creese, A. J., Horner, J. A., Cooper, H. J., High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Comm. in Mass Spectrom.* 2009, 23, 1963-1969.
- [322] Shvartsburg, A. A., Creese, A. J., Smith, R. D., Cooper, H. J., Separation of peptide isomers with variant modified sites by high-resolution differential ion mobility spectrometry. *Anal. Chem.* 2010, 82, 8327-8334.
- [323] Elias, J. E., Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 2007, 4, 207-214.
- [324] Courcelles, M., Lemieux, S., Voisin, L., Meloche, S., Thibault, P., ProteoConnections: a bioinformatics platform to facilitate proteome and phosphoproteome analyses. *Proteomics* 2011.
- [325] Kearney, P., Thibault, P., Bioinformatics meets proteomics--bridging the gap between mass spectrometry data analysis and cell biology. *Journal of bioinformatics and computational biology* 2003, 1, 183-200.
- [326] Bendall, S. C., Hughes, C., Campbell, J. L., Stewart, M. H., *et al.*, An enhanced mass spectrometry approach reveals human embryonic stem cell growth factors in culture. *Mol Cell Proteomics* 2009, 8, 421-432.
- [327] Bonneil, E., Brunet, S., Jaquinod, M., Hui, J. P. M., *et al.*, in: Ivanov, A. R., Lazarev, A. V. (Eds.), *Sample Preparation in Biological Mass Spectrometry*, Springer Science in press.
- [328] Singer, D., Kuhlmann, J., Muschket, M., Hoffmann, R., Separation of multiphosphorylated peptide isomers by hydrophilic interaction chromatography on an aminopropyl phase. *Anal. Chem.* 2010, 82, 6409-6414.
- [329] Kim, S. E., Cho, J. Y., Kim, K. S., Lee, S. J., *et al.*, Drosophila PI3 kinase and Akt involved in insulin-stimulated proliferation and ERK pathway activation in Schneider cells. *Cell Signal* 2004, 16, 1309-1317.
- [330] Taniguchi, C. M., Kondo, T., Sajan, M., Luo, J., *et al.*, Divergent regulation of hepatic glucose and lipid metabolism by phosphoinositide 3-kinase via Akt and PKC λ /zeta. *Cell Metab* 2006, 3, 343-353.
- [331] Kular, G., Loubtchenkov, M., Swigart, P., Whatmore, J., *et al.*, Co-operation of phosphatidylinositol transfer protein with phosphoinositide 3-kinase gamma in the formylmethionyl-leucylphenylalanine-dependent production of phosphatidylinositol 3,4,5-trisphosphate in human neutrophils. *Biochem J* 1997, 325 (Pt 2), 299-301.
- [332] Xu, A. W., Kaelin, C. B., Takeda, K., Akira, S., *et al.*, PI3K integrates the action of insulin and leptin on hypothalamic neurons. *J Clin Invest* 2005, 115, 951-958.
- [333] Britton, J. S., Lockwood, W. K., Li, L., Cohen, S. M., Edgar, B. A., Drosophila's insulin/PI3-kinase pathway coordinates cellular metabolism with nutritional conditions. *Dev Cell* 2002, 2, 239-249.

- [334] Frias, M. A., Thoreen, C. C., Jaffe, J. D., Schroder, W., *et al.*, mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s. *Curr Biol* 2006, 16, 1865-1870.
- [335] Prescott, G. R., Jenkins, R. E., Walsh, C. M., Morgan, A., Phosphorylation of cysteine string protein on Serine 10 triggers 14-3-3 protein binding. *Biochem Biophys Res Commun* 2008, 377, 809-814.
- [336] Donkor, J., Sariahmetoglu, M., Dewald, J., Brindley, D. N., Reue, K., Three mammalian lipins act as phosphatidate phosphatases with distinct tissue expression patterns. *J Biol Chem* 2007, 282, 3450-3457.
- [337] Huffman, T. A., Mothe-Satney, I., Lawrence, J. C., Jr., Insulin-stimulated phosphorylation of lipin mediated by the mammalian target of rapamycin. *Proc Natl Acad Sci U S A* 2002, 99, 1047-1052.
- [338] Peterfy, M., Harris, T. E., Fujita, N., Reue, K., Insulin-stimulated interaction with 14-3-3 promotes cytoplasmic localization of lipin-1 in adipocytes. *J Biol Chem* 2010, 285, 3857-3864.
- [339] Kim, M., Jung, J., Lee, K., Roles of ERK, PI3 kinase, and PLC-gamma pathways induced by overexpression of translationally controlled tumor protein in HeLa cells. *Arch Biochem Biophys* 2009, 485, 82-87.
- [340] Yang, J. T., Laymon, R. A., Goldstein, L. S., A three-domain structure of kinesin heavy chain revealed by DNA sequence and microtubule binding analyses. *Cell* 1989, 56, 879-889.
- [341] Sossey-Alaoui, K., Head, K., Nowak, N., Cowell, J. K., Genomic organization and expression profile of the human and mouse WAVE gene family. *Mamm Genome* 2003, 14, 314-322.
- [342] Welsh, G. I., Stokes, C. M., Wang, X., Sakaue, H., *et al.*, Activation of translation initiation factor eIF2B by insulin requires phosphatidylinositol 3-kinase. *FEBS Lett* 1997, 410, 418-422.
- [343] Kim, S. J., Kahn, C. R., Insulin regulation of mitogen-activated protein kinase kinase (MEK), mitogen-activated protein kinase and casein kinase in the cell nucleus: a possible role in the regulation of gene expression. *Biochem J* 1997, 323 (Pt 3), 621-627.
- [344] Hunter, T., The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell* 2007, 28, 730-738.
- [345] Yao, Q., Li, H., Liu, B. Q., Huang, X. Y., Guo, L., SUMOylation-regulated protein phosphorylation, evidence from quantitative phosphoproteomics analyses. *J Biol Chem* 2011, 286, 27342-27349.
- [346] Guevremont, R., High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *J Chromatogr A* 2004, 1058, 3-19.
- [347] Rodriguez-Viciana, P., Warne, P. H., Vanhaesebroeck, B., Waterfield, M. D., Downward, J., Activation of phosphoinositide 3-kinase by interaction with Ras and by point mutation. *EMBO J* 1996, 15, 2442-2451.
- [348] Franke, T. F., PI3K/Akt: getting it right matters. *Oncogene* 2008, 27, 6473-6488.
- [349] Yang, X., Friedman, A., Nagpal, S., Perrimon, N., Asara, J. M., Use of a label-free quantitative platform based on MS/MS average TIC to calculate dynamics of protein complexes in insulin signaling. *J Biomol Tech* 2009, 20, 272-277.

- [350] Kulkarni, M. M., Booker, M., Silver, S. J., Friedman, A., *et al.*, Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat Methods* 2006, 3, 833-838.
- [351] Ma, Y., Creanga, A., Lum, L., Beachy, P. A., Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature* 2006, 443, 359-363.
- [352] Kubota, K., Anjum, R., Yu, Y., Kunz, R. C., *et al.*, Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. *Nat Biotechnol* 2009, 27, 933-940.
- [353] Xue, Y., Li, A., Wang, L., Feng, H., Yao, X., PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics* 2006, 20, 163.
- [354] Orme, M. H., Alrubaie, S., Bradley, G. L., Walker, C. D., Leever, S. J., Input from Ras is required for maximal PI(3)K signalling in *Drosophila*. *Nat Cell Biol* 2006, 8, 1298-1302.
- [355] Zhai, B., Villen, J., Beausoleil, S. A., Mintseris, J., Gygi, S. P., Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J Proteome Res* 2008, 7, 1675-1682.
- [356] Poltilove, R. M., Jacobs, A. R., Haft, C. R., Xu, P., Taylor, S. I., Characterization of *Drosophila* insulin receptor substrate. *J Biol Chem* 2000, 275, 23346-23354.
- [357] Chen, F., Rebay, I., split ends, a new component of the *Drosophila* EGF receptor pathway, regulates development of midline glial cells. *Curr Biol* 2000, 10, 943-946.
- [358] Wiellette, E. L., Harding, K. W., Mace, K. A., Ronshaugen, M. R., *et al.*, spen encodes an RNP motif protein that interacts with Hox pathways to repress the development of head-like sclerites in the *Drosophila* trunk. *Development* 1999, 126, 5373-5385.
- [359] Ackerman, P., Glover, C. V., Osheroff, N., Phosphorylation of DNA topoisomerase II by casein kinase II: modulation of eukaryotic topoisomerase II activity in vitro. *Proc Natl Acad Sci U S A* 1985, 82, 3164-3168.
- [360] Corbett, A. H., DeVore, R. F., Osheroff, N., Effect of casein kinase II-mediated phosphorylation on the catalytic cycle of topoisomerase II. Regulation of enzyme activity by enhancement of ATP hydrolysis. *J Biol Chem* 1992, 267, 20513-20518.
- [361] Daum, J. R., Gorbsky, G. J., Casein kinase II catalyzes a mitotic phosphorylation on threonine 1342 of human DNA topoisomerase IIalpha, which is recognized by the 3F3/2 phosphoepitope antibody. *J Biol Chem* 1998, 273, 30622-30629.
- [362] Escargueil, A. E., Plisov, S. Y., Filhol, O., Cochet, C., Larsen, A. K., Mitotic phosphorylation of DNA topoisomerase II alpha by protein kinase CK2 creates the MPM-2 phosphoepitope on Ser-1469. *J Biol Chem* 2000, 275, 34710-34718.
- [363] Wells, N. J., Addison, C. M., Fry, A. M., Ganapathi, R., Hickson, I. D., Serine 1524 is a major site of phosphorylation on human topoisomerase II alpha protein in vivo and is a substrate for casein kinase II in vitro. *J Biol Chem* 1994, 269, 29746-29751.
- [364] Yue, L., Spradling, A. C., hu-li tai shao, a gene required for ring canal formation during *Drosophila* oogenesis, encodes a homolog of adducin. *Genes Dev* 1992, 6, 2443-2454.

- [365] Ashley, C. T., Pendleton, C. G., Jennings, W. W., Saxena, A., Glover, C. V., Isolation and sequencing of cDNA clones encoding *Drosophila* chromosomal protein D1. A repeating motif in proteins which recognize at DNA. *J Biol Chem* 1989, 264, 8394-8401.
- [366] Buszczak, M., Paterno, S., Spradling, A. C., *Drosophila* stem cells share a common requirement for the histone H2B ubiquitin protease scrawny. *Science* 2009, 323, 248-251.
- [367] Karpac, J., Younger, A., Jasper, H., Dynamic coordination of innate immune signaling and insulin signaling regulates systemic responses to localized DNA damage. *Dev Cell* 2011, 20, 841-854.
- [368] Schneider, C. C., Hessenauer, A., Gotz, C., Montenarh, M., DMAT, an inhibitor of protein kinase CK2 induces reactive oxygen species and DNA double strand breaks. *Oncol Rep* 2009, 21, 1593-1597.
- [369] Zwicker, F., Ebert, M., Huber, P. E., Debus, J., Weber, K. J., A specific inhibitor of protein kinase CK2 delays gamma-H2Ax foci removal and reduces clonogenic survival of irradiated mammalian cells. *Radiat Oncol* 2011, 6, 15.
- [370] Loizou, J. I., El-Khamisy, S. F., Zlatanou, A., Moore, D. J., *et al.*, The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks. *Cell* 2004, 117, 17-28.
- [371] Li, D., Meier, U. T., Dobrowolska, G., Krebs, E. G., Specific interaction between casein kinase 2 and the nucleolar protein Nopp140. *J Biol Chem* 1997, 272, 3773-3779.
- [372] Pinna, L. A., Casein kinase 2: an 'eminence grise' in cellular regulation? *Biochim Biophys Acta* 1990, 1054, 267-284.
- [373] Roach, P. J., Multisite and hierarchal protein phosphorylation. *J Biol Chem* 1991, 266, 14139-14142.
- [374] Song, C., Li, Z., Erbe, A. K., Savic, A., Dovat, S., Regulation of Ikaros function by casein kinase 2 and protein phosphatase 1. *World J Biol Chem* 2011, 2, 126-131.
- [375] Thingholm, T. E., Jensen, O. N., Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* 2009, 9, 1451-1468.
- [376] Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., *et al.*, Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nat Methods* 2011, 8, 821-827.
- [377] Trost, M., English, L., Lemieux, S., Courcelles, M., *et al.*, The phagosomal proteome in interferon-gamma-activated macrophages. *Immunity* 2009, 30, 143-154.
- [378] Otvos, L., Jr., Tangoren, I. A., Wroblewski, K., Hollosi, M., Lee, V. M., Reversed-phase high-performance liquid chromatographic separation of synthetic phosphopeptide isomers. *J Chromatogr* 1990, 512, 265-272.
- [379] Hoffmann, R., Segal, M., Otvos, L., Separation of sets of mono- and diphosphorylated peptides by reversed-phase high performance liquid chromatography. *Analytica Chimica Acta* 1997, 352, 327-333.
- [380] Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., *et al.*, Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005, 4, 310-327.

- [381] Ballif, B. A., Roux, P. P., Gerber, S. A., MacKeigan, J. P., *et al.*, Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc Natl Acad Sci U S A* 2005, 102, 667-672.
- [382] Boeri Erba, E., Matthiesen, R., Bunkenborg, J., Schulze, W. X., *et al.*, Quantitation of multisite EGF receptor phosphorylation using mass spectrometry and a novel normalization approach. *J Proteome Res* 2007, 6, 2768-2785.
- [383] Cunningham, D. L., Sweet, S. M., Cooper, H. J., Heath, J. K., Differential phosphoproteomics of fibroblast growth factor signaling: identification of Src family kinase-mediated phosphorylation events. *J Proteome Res* 2010, 9, 2317-2328.
- [384] Sweet, S. M., Mardakheh, F. K., Ryan, K. J., Langton, A. J., *et al.*, Targeted online liquid chromatography electron capture dissociation mass spectrometry for the localization of sites of in vivo phosphorylation in human Sprouty2. *Anal Chem* 2008, 80, 6650-6657.
- [385] Tholey, A., Toll, H., Huber, C. G., Separation and detection of phosphorylated and nonphosphorylated peptides in liquid chromatography-mass spectrometry using monolithic columns and acidic or alkaline mobile phases. *Anal Chem* 2005, 77, 4618-4625.
- [386] Singer, D., Kuhlmann, J., Muschket, M., Hoffmann, R., Separation of multiphosphorylated peptide isomers by hydrophilic interaction chromatography on an aminopropyl phase. *Anal Chem* 2010, 82, 6409-6414.
- [387] Muetzelburg, M. V., Hoffmann, R., Separation of multiphosphorylated peptide isomers by CZE. *Electrophoresis* 2008, 29, 4381-4385.
- [388] Bridon, G., Bonnel, E., Muratore-Schroeder, T., Caron-Lizotte, O., Thibault, P., Improvement of phosphoproteome analyses using FAIMS and decision tree fragmentation; Application to the insulin signalling pathway in *Drosophila melanogaster* S2 cells. *J Proteome Res* 2011.
- [389] Xuan, Y., Creese, A. J., Horner, J. A., Cooper, H. J., High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Commun Mass Spectrom* 2009, 23, 1963-1969.
- [390] Perlova, T. Y., Goloborodko, A. A., Margolin, Y., Pridatchenko, M. L., *et al.*, Retention time prediction using the model of liquid chromatography of biomacromolecules at critical conditions in LC-MS phosphopeptide analysis. *Proteomics* 2010, 10, 3458-3468.
- [391] Kyte, J., Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982, 157, 105-132.
- [392] Winter, D., Pipkorn, R., Lehmann, W. D., Separation of peptide isomers and conformers by ultra performance liquid chromatography. *J Sep Sci* 2009, 32, 1111-1119.
- [393] Krokhin, O. V., Antonovici, M., Ens, W., Wilkins, J. A., Standing, K. G., Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* 2006, 78, 6645-6650.
- [394] Kelstrup, C. D., Hekmat, O., Francavilla, C., Olsen, J. V., Pinpointing phosphorylation sites: Quantitative filtering and a novel site-specific x-ion fragment. *J Proteome Res* 2011, 10, 2937-2948.

- [395] Bose, R., Holbert, M. A., Pickin, K. A., Cole, P. A., Protein tyrosine kinase-substrate interactions. *Current opinion in structural biology* 2006, 16, 668-675.
- [396] Gnad, F., Gunawardena, J., Mann, M., PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res* 2011, 39, D253-260.
- [397] Ghosh, G., Adams, J. A., Phosphorylation mechanism and structure of serine-arginine protein kinases. *The FEBS journal*, 278, 587-597.
- [398] Li, H., Unveiling substrate RNA binding to H/ACA RNPs: one side fits all. *Current opinion in structural biology* 2008, 18, 78-85.
- [399] Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M., *et al.*, NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* 2008, 36, D695-699.
- [400] Tsakiridis, T., Wang, Q., Taha, C., Grinstein, S., *et al.*, Involvement of the actin network in insulin signalling. *Soc Gen Physiol Ser* 1997, 52, 257-271.
- [401] Pessin, J. E., Thurmond, D. C., Elmendorf, J. S., Coker, K. J., Okada, S., Molecular basis of insulin-stimulated GLUT4 vesicle trafficking. Location! Location! Location! *J Biol Chem* 1999, 274, 2593-2596.
- [402] Tawfic, S., Yu, S., Wang, H., Faust, R., *et al.*, Protein kinase CK2 signal in neoplasia. *Histology and histopathology* 2001, 16, 573-582.
- [403] Guerra, B., Protein kinase CK2 subunits are positive regulators of AKT kinase. *Int J Oncol* 2006, 28, 685-693.
- [404] Chen, M., Li, D., Krebs, E. G., Cooper, J. A., The casein kinase II beta subunit binds to Mos and inhibits Mos activity. *Mol Cell Biol* 1997, 17, 1904-1912.
- [405] Guerra, B., Issinger, O. G., Wang, J. Y., Modulation of human checkpoint kinase Chk1 by the regulatory beta-subunit of protein kinase CK2. *Oncogene* 2003, 22, 4933-4942.
- [406] Meng, R., Al-Quobaili, F., Muller, I., Gotz, C., *et al.*, CK2 phosphorylation of Pdx-1 regulates its transcription factor activity. *Cell Mol Life Sci* 2010, 67, 2481-2489.
- [407] Al-Khouri, A. M., Ma, Y., Togo, S. H., Williams, S., Mustelin, T., Cooperative phosphorylation of the tumor suppressor phosphatase and tensin homologue (PTEN) by casein kinases and glycogen synthase kinase 3beta. *J Biol Chem* 2005, 280, 35195-35202.
- [408] Keller, D. M., Zeng, X., Wang, Y., Zhang, Q. H., *et al.*, A DNA damage-induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1. *Mol Cell* 2001, 7, 283-292.
- [409] Kato, T., Jr., Delhase, M., Hoffmann, A., Karin, M., CK2 Is a C-Terminal I kappaB Kinase Responsible for NF-kappaB Activation during the UV Response. *Mol Cell* 2003, 12, 829-839.
- [410] Krohn, N. M., Stemmer, C., Fojan, P., Grimm, R., Grasser, K. D., Protein kinase CK2 phosphorylates the high mobility group domain protein SSRP1, inducing the recognition of UV-damaged DNA. *J Biol Chem* 2003, 278, 12710-12715.

[411] Radimerski, T., Mini, T., Schneider, U., Wettenhall, R. E., *et al.*, Identification of insulin-induced sites of ribosomal protein S6 phosphorylation in *Drosophila melanogaster*. *Biochemistry* 2000, 39, 5766-5774.

**Annexe 1 : Figures et tableaux supplémentaires du
Chapitre 1**

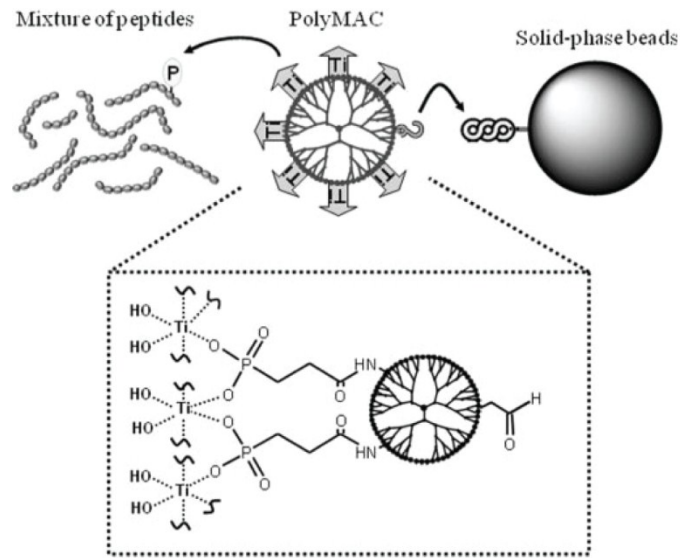


Figure A1. 1. Schéma de la stratégie PolyMAC pour l'enrichissement de phosphopeptides.

La méthode consiste en deux étapes soit la chélation homogène entre les phosphopeptides et le réactif PolyMAC-Ti en solution puis la réaction hétérogène entre le réactif PolyMAC et le support en phase solide. Reproduction de [179] avec la permission de The American Society for Biochemistry and Molecular Biology, Molecular and Cellular Proteomics, copyright 2010.

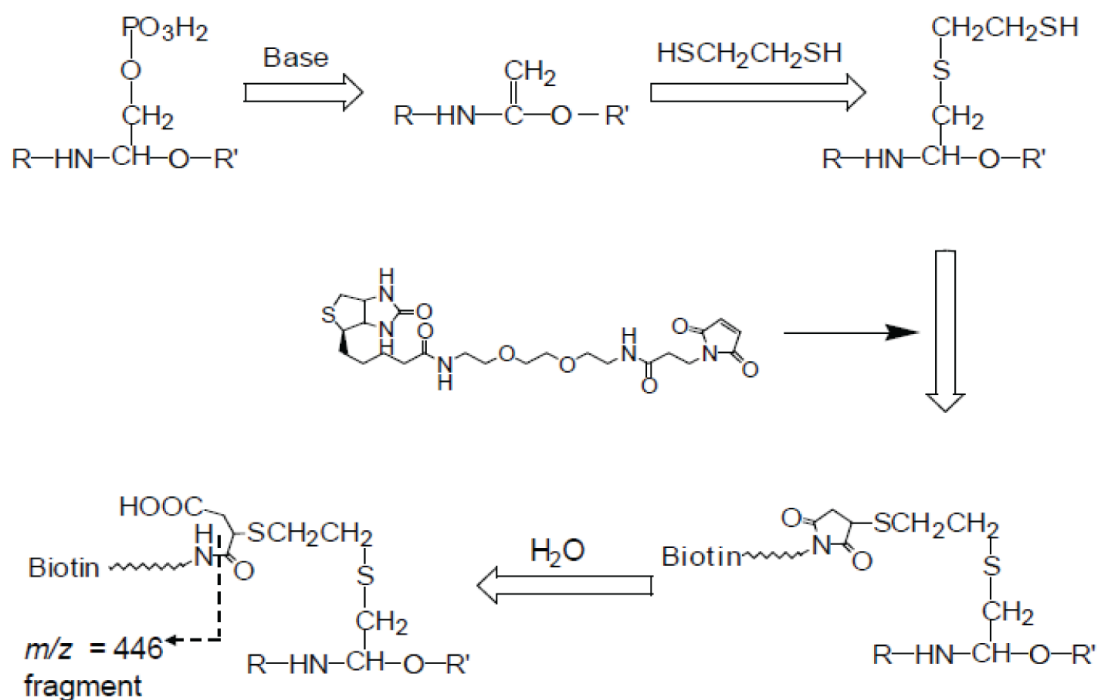


Figure A1. 2. Schéma de la conversion chimique d'une phosphosérine en un résidu biotinylé en vue de sa purification.

Suite à l'élimination β du groupement phosphate en conditions basiques, un groupement sulfhydryle est ajouté par addition de type Michael permettant ensuite la liaison d'un groupe biotine pour sa purification subséquente. La même stratégie d'enrichissement peut être appliquée aux phosphothréonines. Adaptation de [190] avec la permission de Macmillan Publishers Ltd, Nature Biotechnology, copyright 2001.

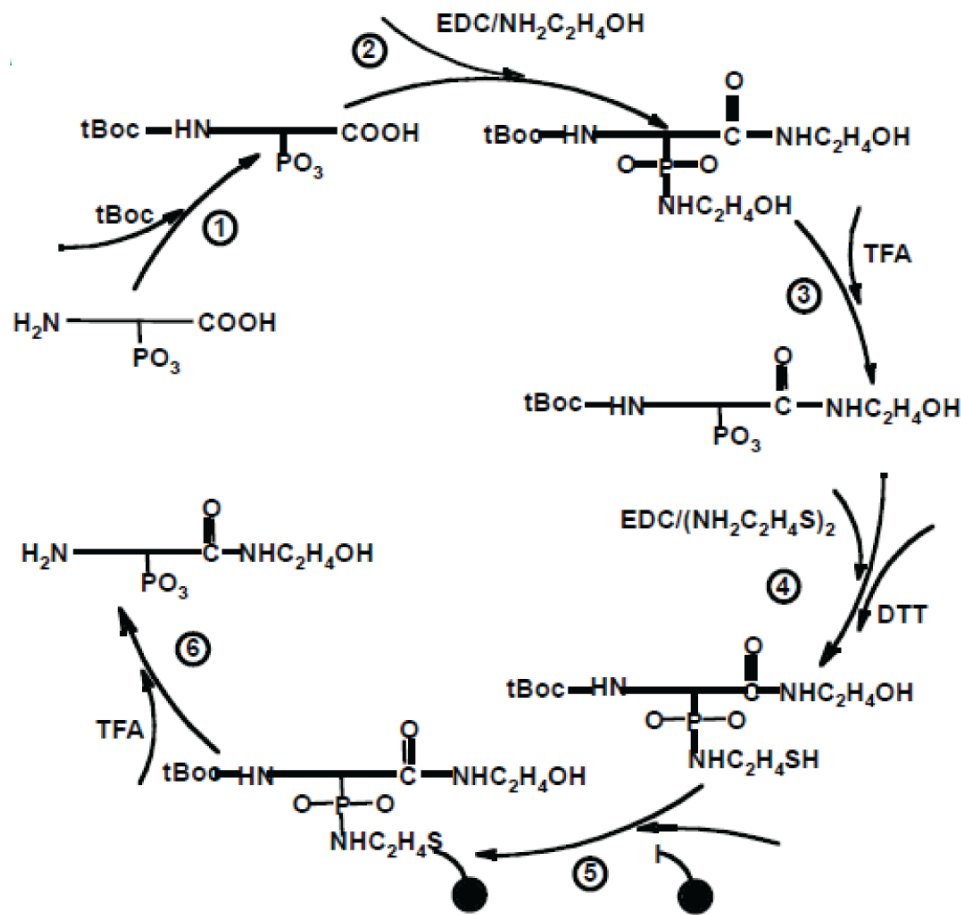


Figure A1. 3. Séquence de six réactions chimiques menant à l'isolation sélective de phosphopeptides dans une mixture complexe par chimie de phosphoramidate.

1) La première étape consiste en la protection des amines par le t-butyl-dicarbonat (tBoc). 2) La réaction de condensation entre les peptides et les amines en excès est catalysée par le carbodiimide afin de former des liens amide et phosphoramidate respectivement sur les groupements carboxylate et phosphate. 3) Le groupement phosphate est régénéré par hydrolyse acide de la liaison phosphoramidate. 4) Une réaction de condensation catalysée par le carbodiimide attache une cystamine au groupement phosphate régénéré. La réduction du groupement disulfide interne de la cystamine génère un groupe sulfhydryle pour chaque phosphate. 5) Capture sur phase solide : les excès de cystamine, agent réducteur et autres sous-produits sont éliminés suite au lavage des peptides retenus sur une résine en phase inverse. 6) Les phosphopeptides sont élués suite au clivage de la liaison phosphoramidate grâce au TFA ce qui va également éliminer le groupement tBoc de protection. Reproduction de [193] avec la permission de Macmillan Publishers Ltd, Nature Biotechnology, copyright 2001.

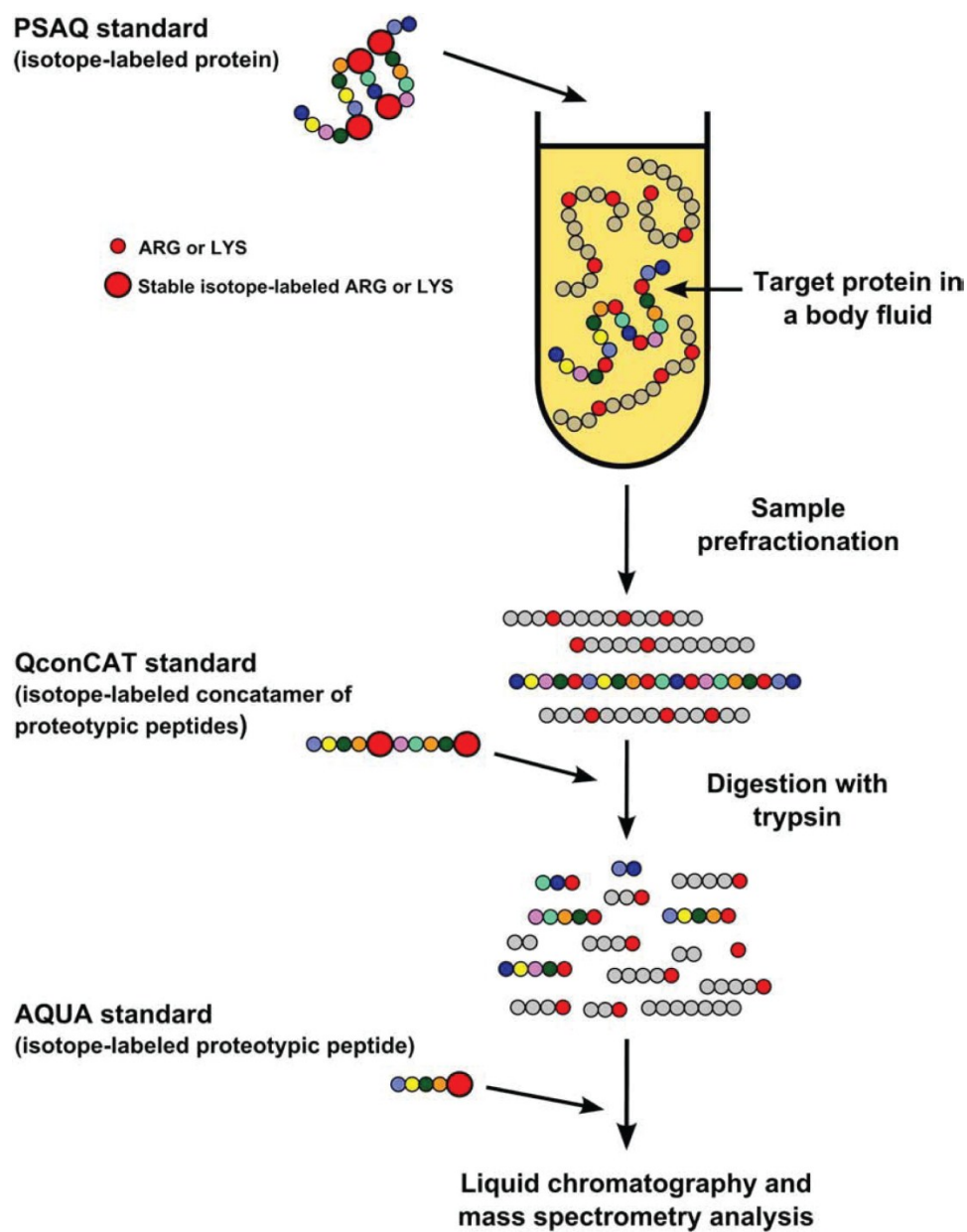


Figure A1. 4. Stratégies utilisant des standards isotopiques pour la quantification absolue de protéines ou de peptides par spectrométrie de masse.

Trois types de standards sont disponibles : PSAQ (Protein Standard Absolute Quantification), QconCAT (Quantification concatamer) et AQUA (Absolute Quantification). Reproduction de [252] avec la permission de Elsevier, Journal of Proteomics, copyright 2009.

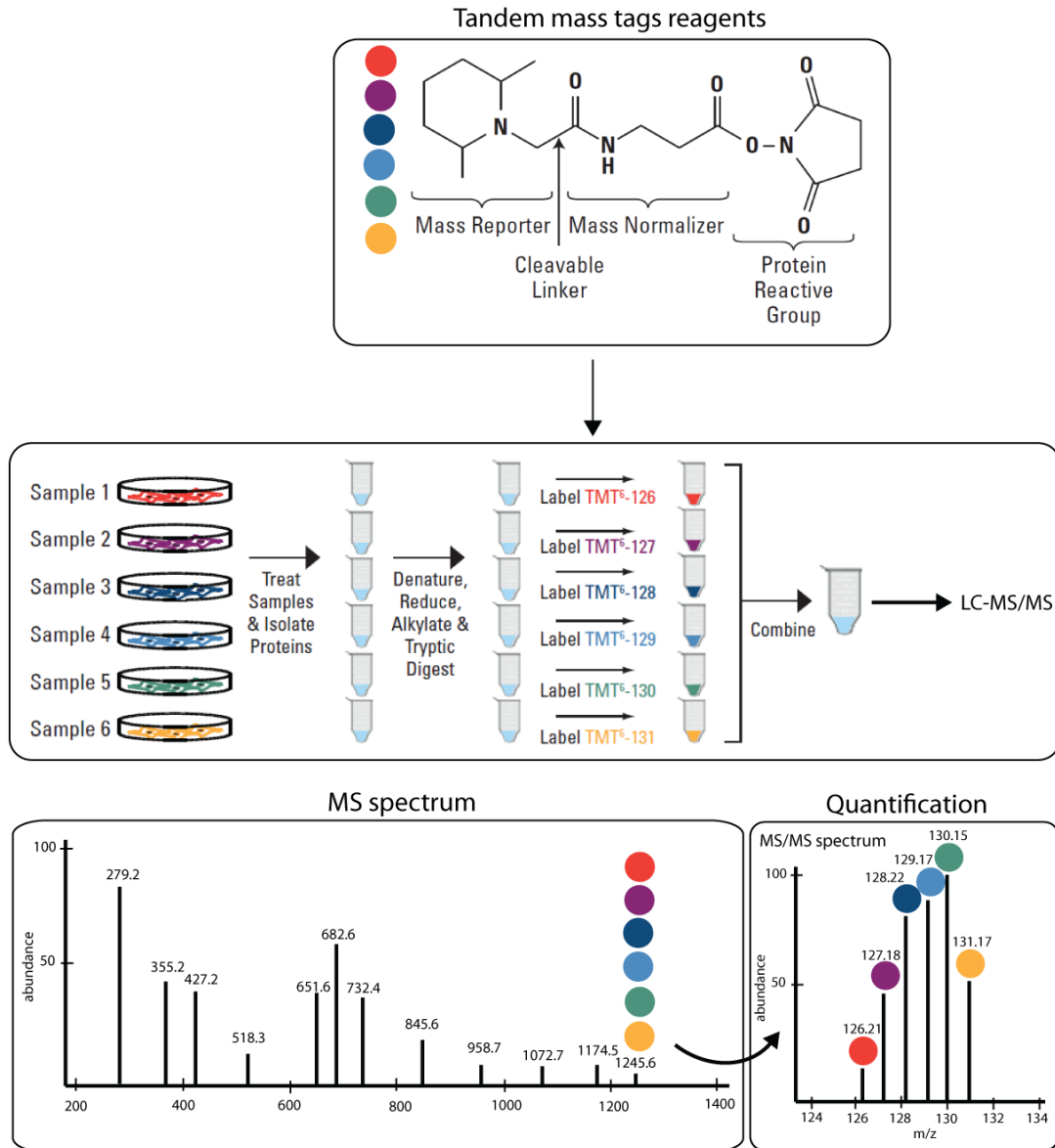


Figure A1. 5. Schéma du protocole opératoire réalisé lors d'un marquage par TMT (Tandem Mass Tags) pour la quantification relative de peptides par spectrométrie de masse.

Chaque reporteur a une masse unique qui permet de le distinguer suite à la fragmentation. Jusqu'à six conditions peuvent être quantifiées et donc comparées simultanément par LC-MS/MS suite au marquage peptidique. La fragmentation par MS/MS du précurseur permet de déterminer l'abondance relative de chaque reporteur qui reflète l'abondance du peptide d'intérêt dans chacune des conditions. Adapté de « Isobaric mass tagging kits », Pierce 2012 (<http://www.piercenet.com/previews/Vol12Iss03/>).

**Annexe 2 : Figures et tableaux supplémentaires du
Chapitre 2**

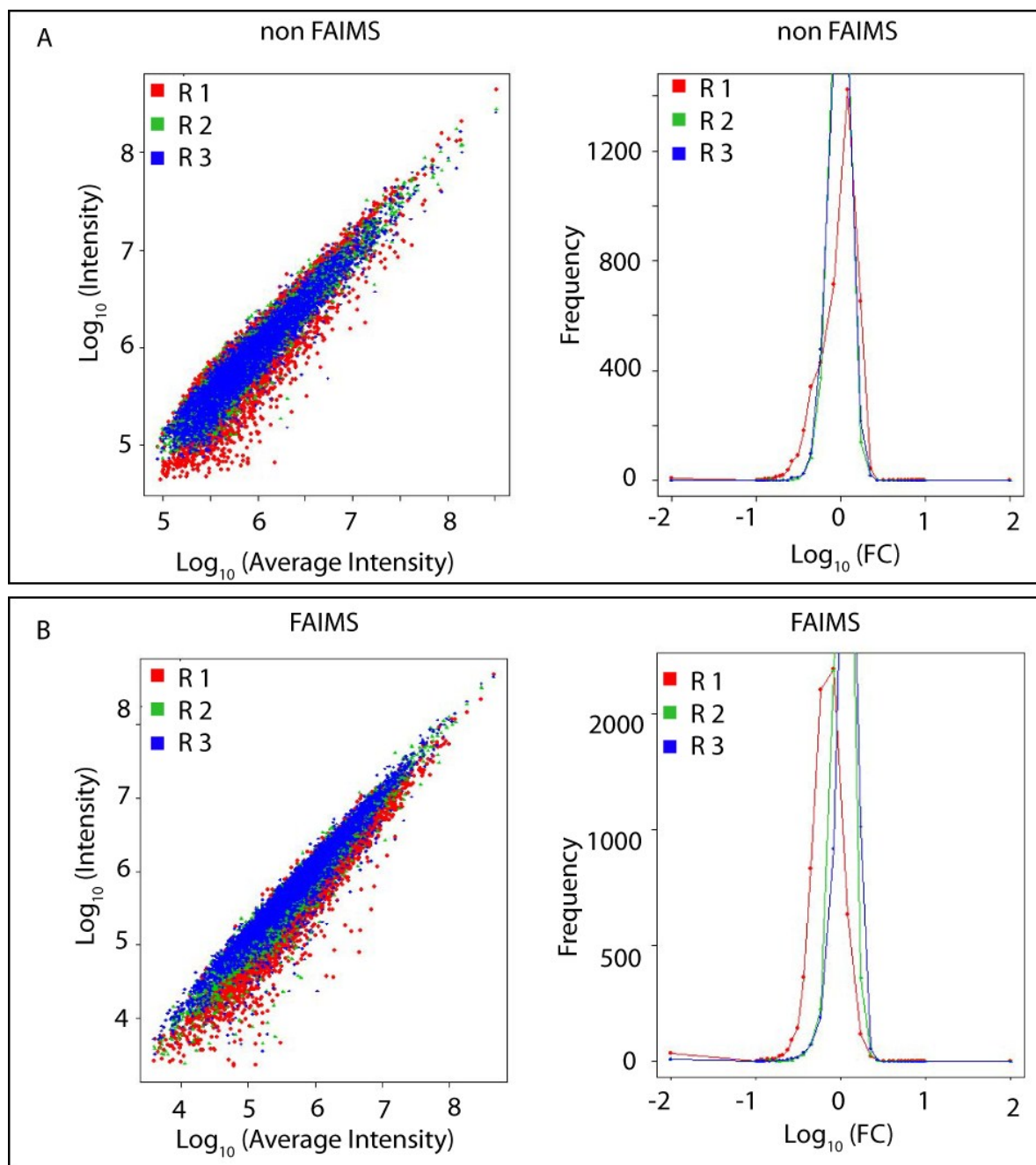


Figure A2. 1. Scatter plots of intensity measurements obtained for 2D-LC-MS/MS experiments performed with and without FAIMS.

2D-LC-MS/MS analyses performed with and without FAIMS yielded RSD values of 45% and 42%, respectively.

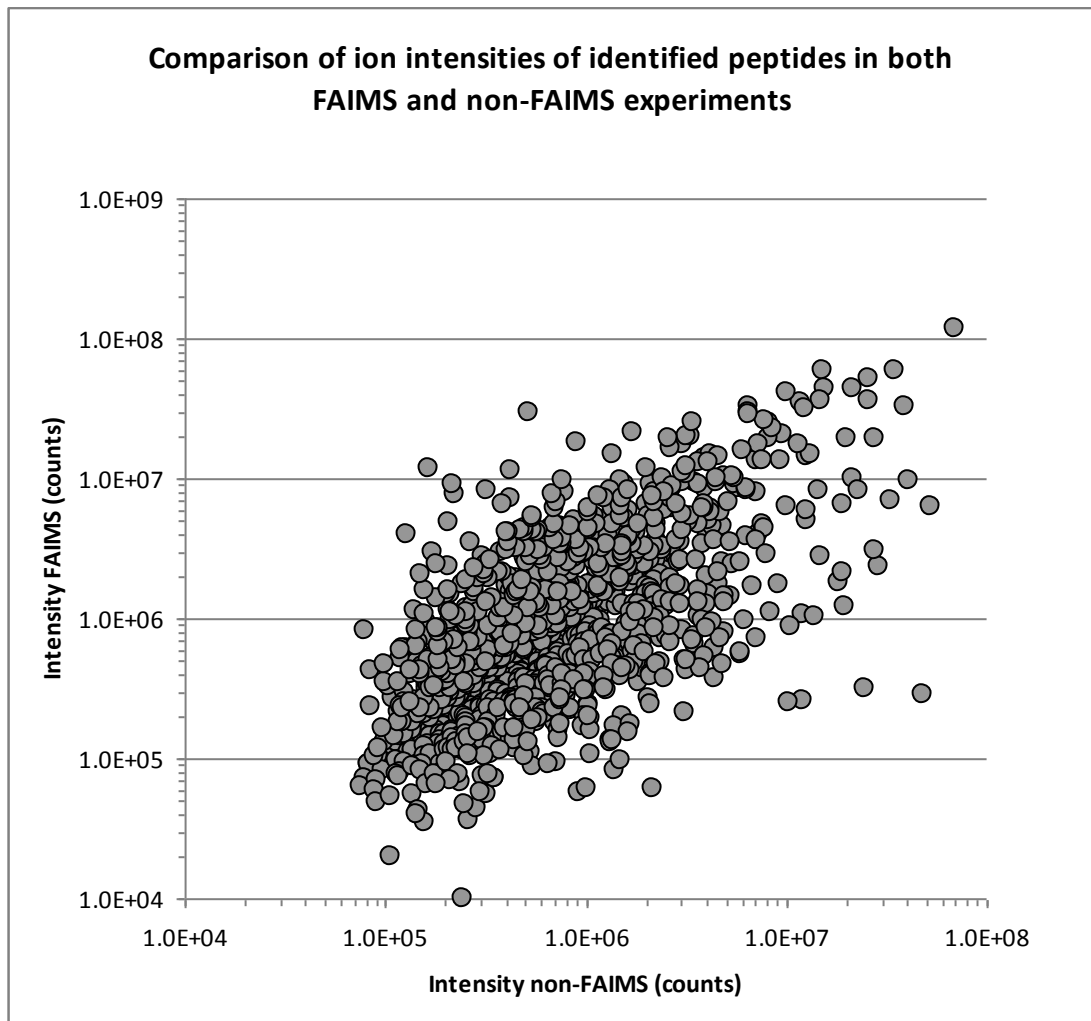
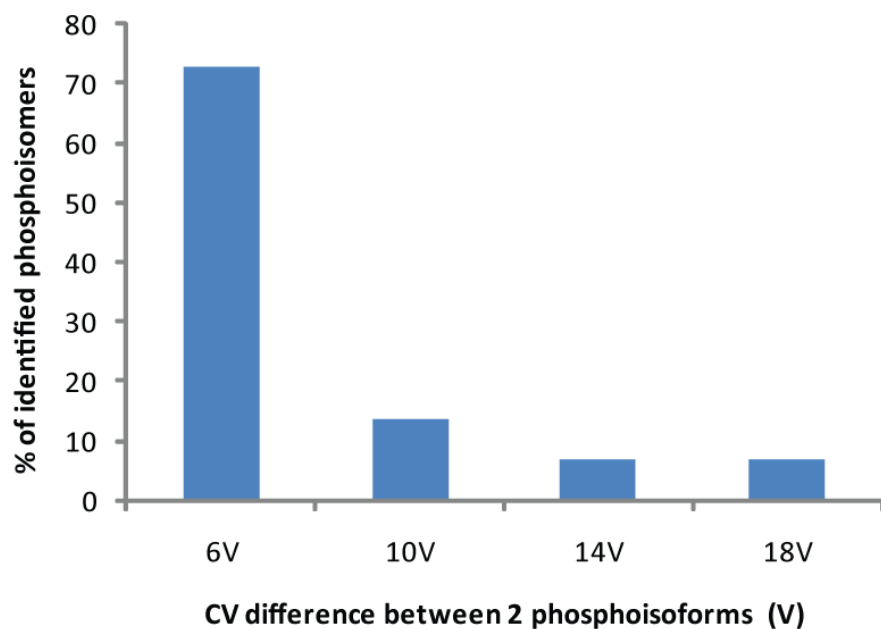


Figure A2. 2. Scatter plot of intensity measurements for peptides identified in both FAIMS and non-FAIMS 2D-LC-MS/MS experiments.

a)



b)

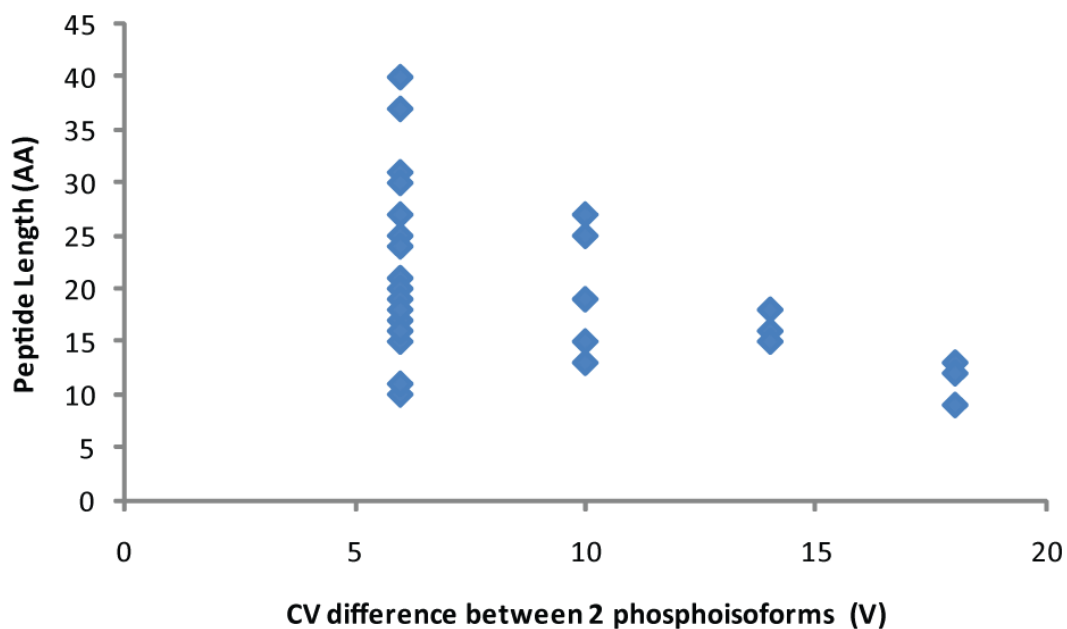


Figure A2. 3. Variation of CV values for phosphopeptides isomers.

(A) Difference of CV for pairs of phosphopeptide isoforms. (B) Distribution of peptide length and difference of CV. Larger differences in CV values were observed for short phosphopeptide isomers.

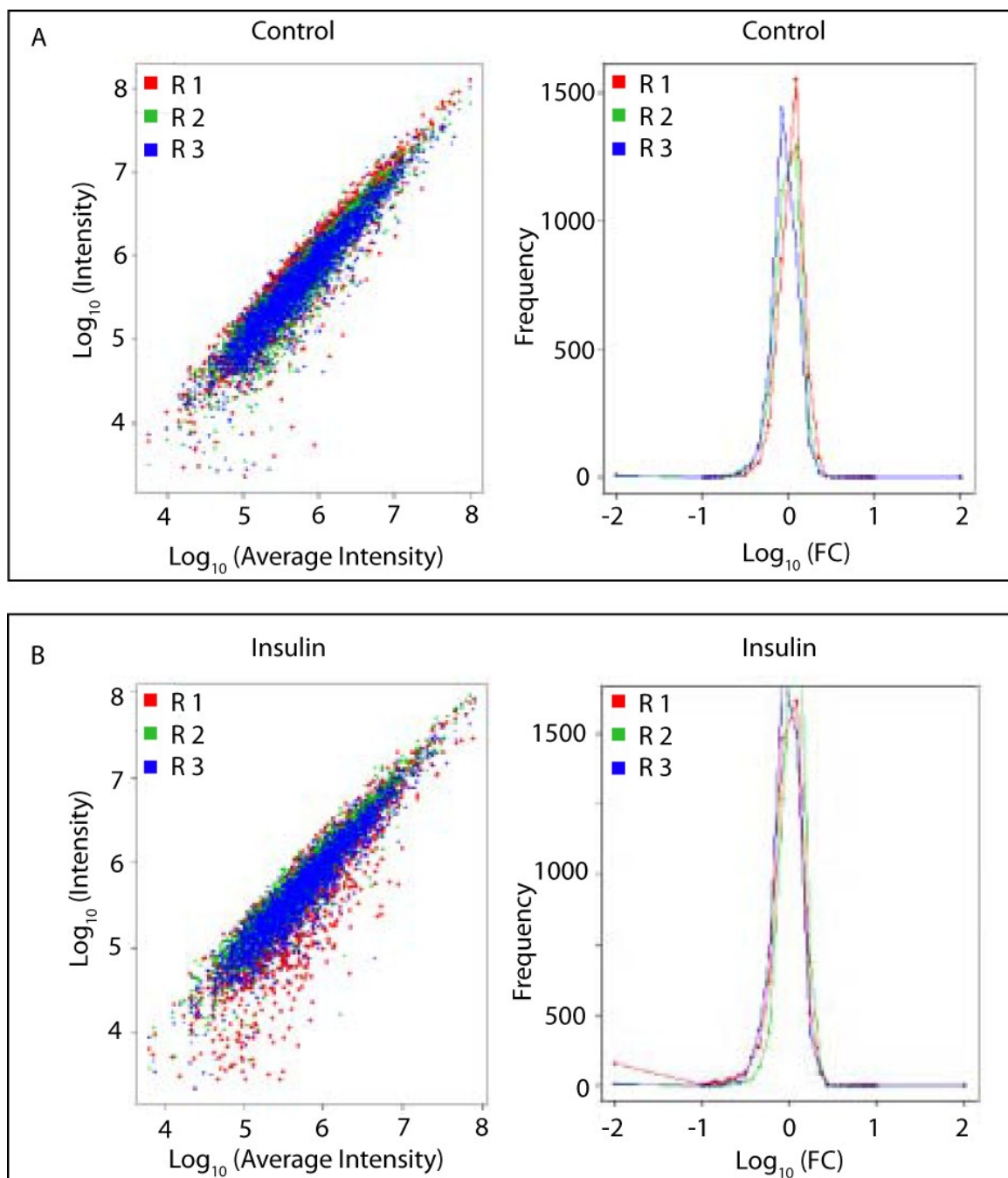


Figure A2. 4. Scatter plot of intensity measurements obtained for 2D-LC-MS/MS analyses of control and insulin treated *D. melanogaster* S2 cells (8 min stimulation, 0.86 μM insulin).

Intensity measurements are compared between three biological replicates of control and insulin-stimulated cells.

Table A2- I. List of all unique phosphosites identified in FAIMS and non-FAIMS experiments conducted in control S2 cells. (CD-ROM).

Table A2- II. List of all unique phosphosites identified in FAIMS experiments in control and insulin-stimulated S2 cells. (CD ROM).

**Annexe 3 : Figures et tableaux supplémentaires du
Chapitre 3**

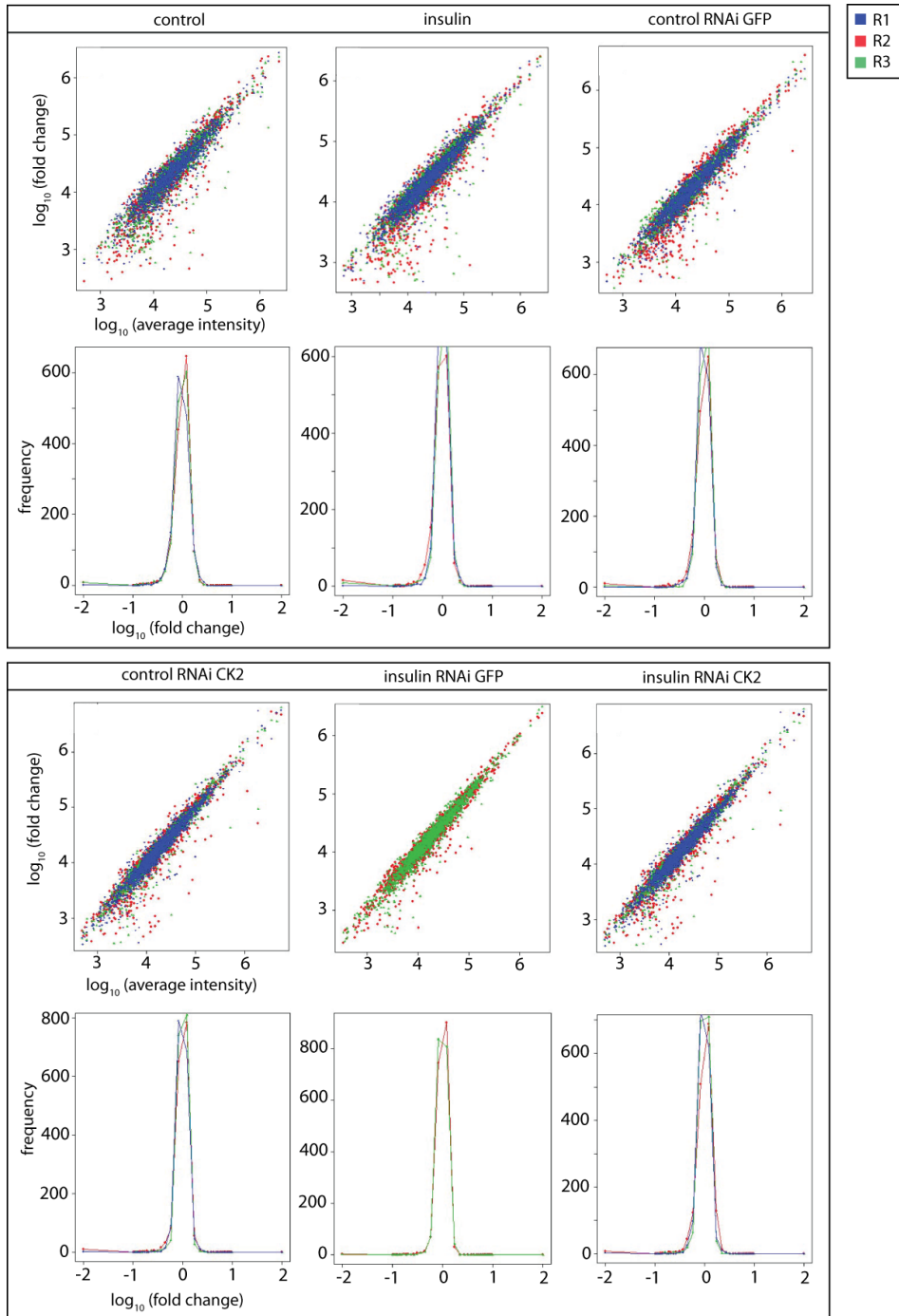


Figure A3. 1. Scatter plot of intensity measurements obtained for FAIMS-1D-LC-MS/MS analyses in Control, Control GFP RNAi, Control CK2 RNAi, Insulin, Insulin GFP RNAi and Insulin CK2 RNAi treated *Drosophila melanogaster* S2 cells.

Peptide intensities are displayed for three biological triplicates (duplicates for Insulin GFP RNAi condition).

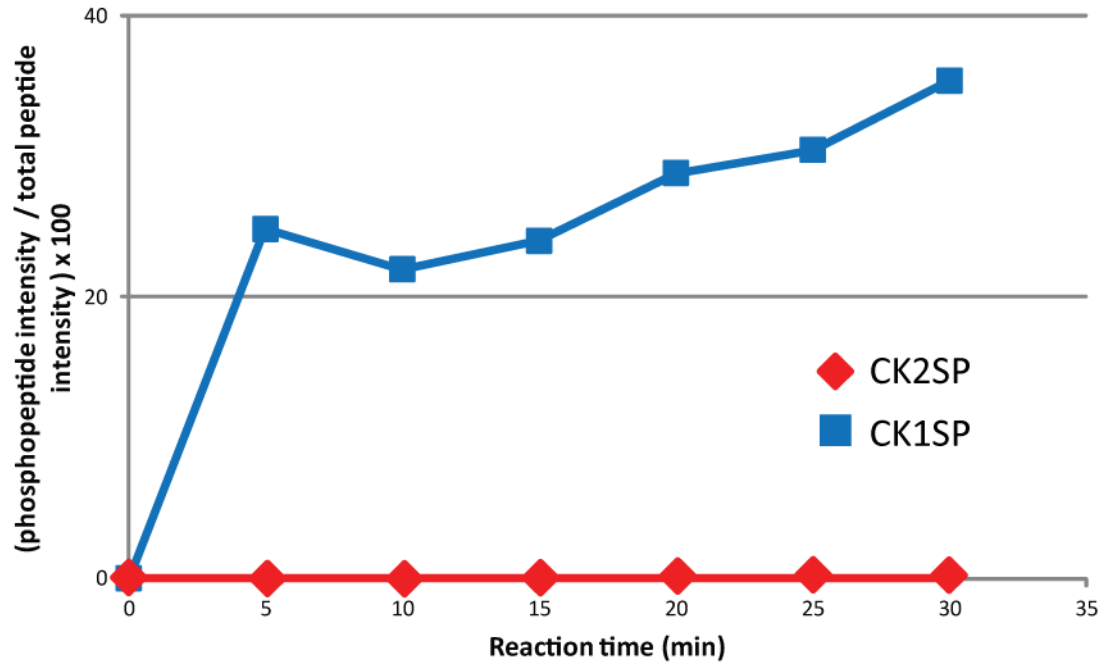


Figure A3. 2. Representation of the ratio of intensity between the phosphorylated form of a specific peptide and the total peptide intensity (phosphorylated + non phosphorylated forms) using CK1 enzyme in *in vitro* kinase assay experiments.

CK1SP (RRKDLHDDEEDEAMSITA) is partially transformed into a phosphorylated form (40%) after 30 min of reaction. CK2SP (negative control) as well as all other peptides do not show any phosphorylation.

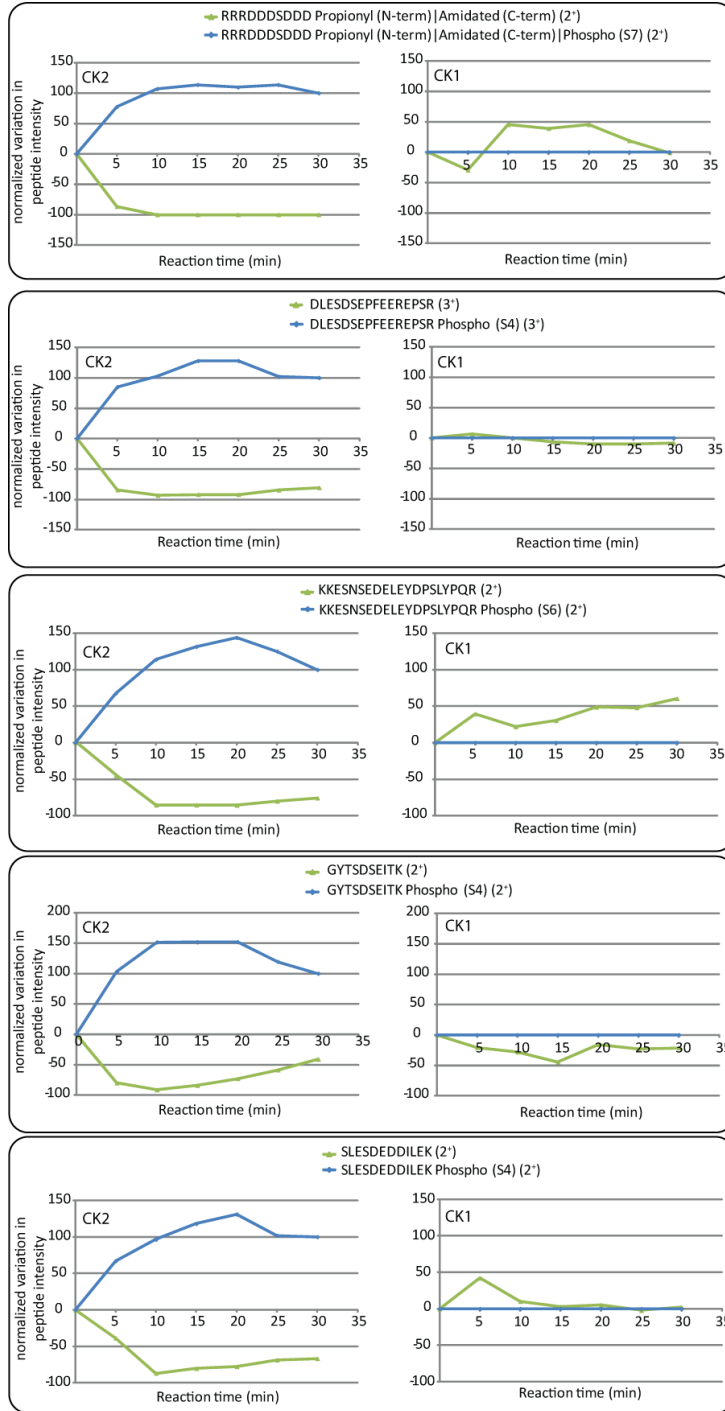


Figure A3.3. Representation of the variation in intensity of 5 non-phosphorylated and phosphorylated peptides using CK2 and CK1 enzymes during *in vitro* kinase assay.

For CK2-specific targets hits, we observe a decrease in intensity for the non-phosphorylated peptide form and an increase in intensity for the phosphorylated counterpart.

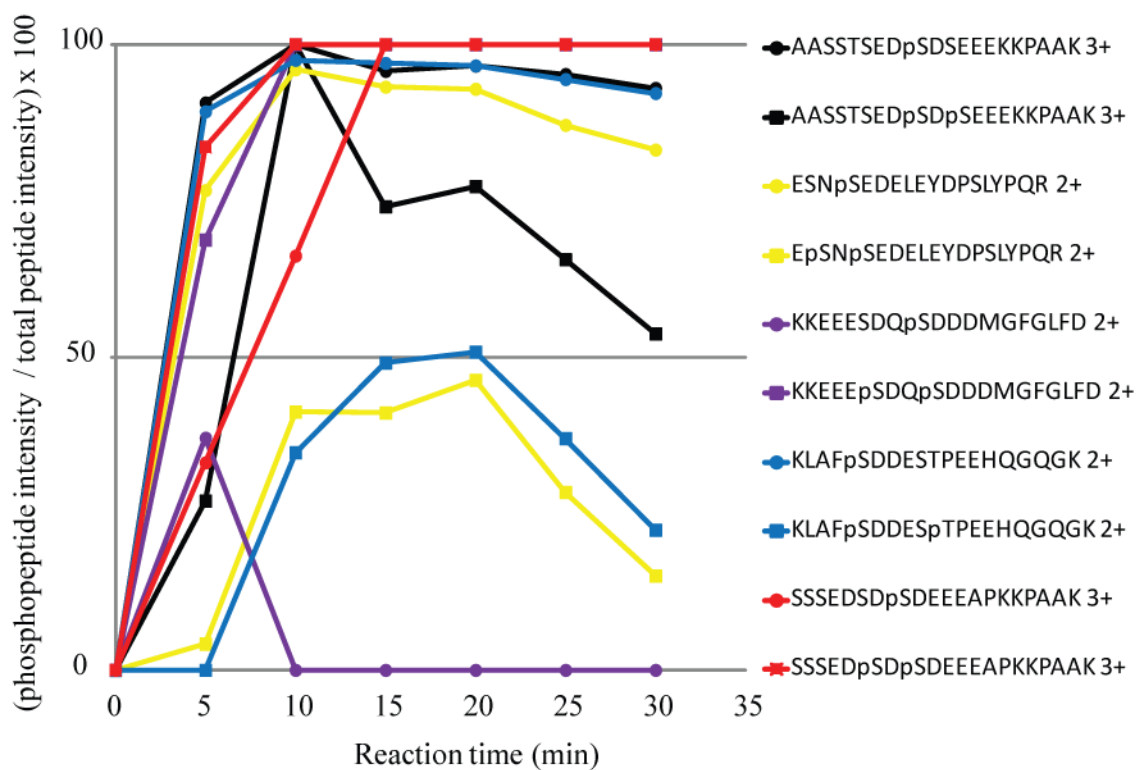


Figure A3. 4. Representation of the phosphorylation kinetic of 5 mono- and multiply-phosphorylated pairs of synthetic peptides using *in vitro* kinase assay and CK2 recombinant enzyme.

The ratio of intensity between the phosphorylated form of a specific peptide and the total peptide intensity (phosphorylated + non phosphorylated forms) for a specific state of charge is displayed. Lines with the same color represent the same peptide sequence, round and square symbols represents the mono- and di-phosphorylated forms, respectively.

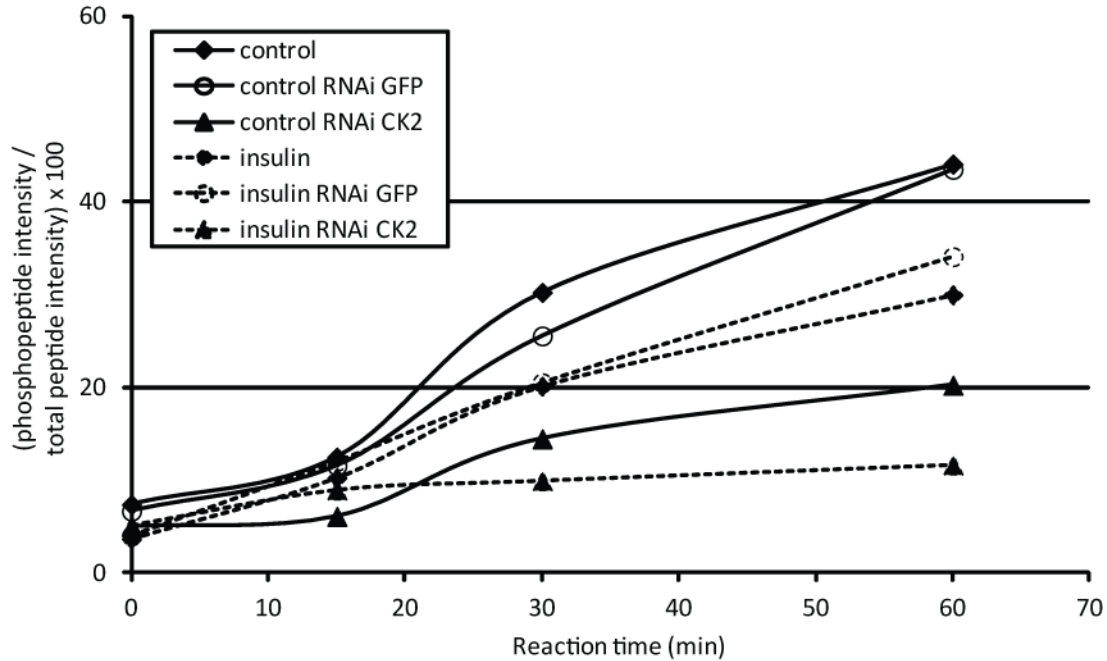


Figure A3. 5. Determination of CK2 activity using *in vitro* kinase assay and CK2SP phosphorylation as a readout.

CK2SP is used as a substrate spiked in Control, Insulin and CK2 RNAi treated S2 cells protein lysates containing endogenous CK2. CK2 is constitutively active and phosphorylates CK2SP in control cells (40% of CK2SP is in a phosphorylated form after 60 min of reaction). In insulin-stimulated cells, only 30% of CK2SP is under a phosphorylated form showing that CK2 is not more active after insulin stimulation. The same trend is observed in GFP RNAi treated cells. A significant decrease in CK2SP phosphorylation is observed in CK2 RNAi treated cells (20% and 10% of CK2SP are detected as phosphorylated in Control CK2 RNAi and Insulin CK2 RNAi S2 cells, respectively). Errors bars are omitted for clarity; average standard deviation is 35% on 3 biological replicates.

Table A3- I. List of phosphosites identified in *Drosophila melanogaster* S2 cells by 1D-LC-FAIMS-MS/MS on a LTQ-Orbitrap Velos. (CD-ROM).

Table A3- II. Complete list of phosphosites affected by insulin or CK2 RNAi treatment showing their corresponding intensity fold change and p-value calculated across 3 biological triplicates. (CD-ROM).

Uniprot ID, CG ID, gene name, peptide sequence and phosphosites position are displayed. CK2-related phosphorylation for *in vitro* kinase assays using synthetic peptides is indicated. High: transformation into a majority phosphorylated form (80-100%) after 30 min of reaction, medium: detection of a phosphorylated peptide (< 50% of the total peptide intensity), no: no phosphorylation detected, N/A: non-tested peptide sequence. The presence of CK1/CK2 consensus motifs are indicated (* represents CK2 known targets according to NetworKIN) as well as phosphosites conservation in mammals. Pink and green backgrounds indicate phosphopeptides up- and down-regulated after insulin stimulation, respectively. Pink and green writings indicate phosphopeptides up- and down-regulated after CK2 RNAi treatment, respectively.

**Annexe 4 : Figures et tableaux supplémentaires du
Chapitre 4**

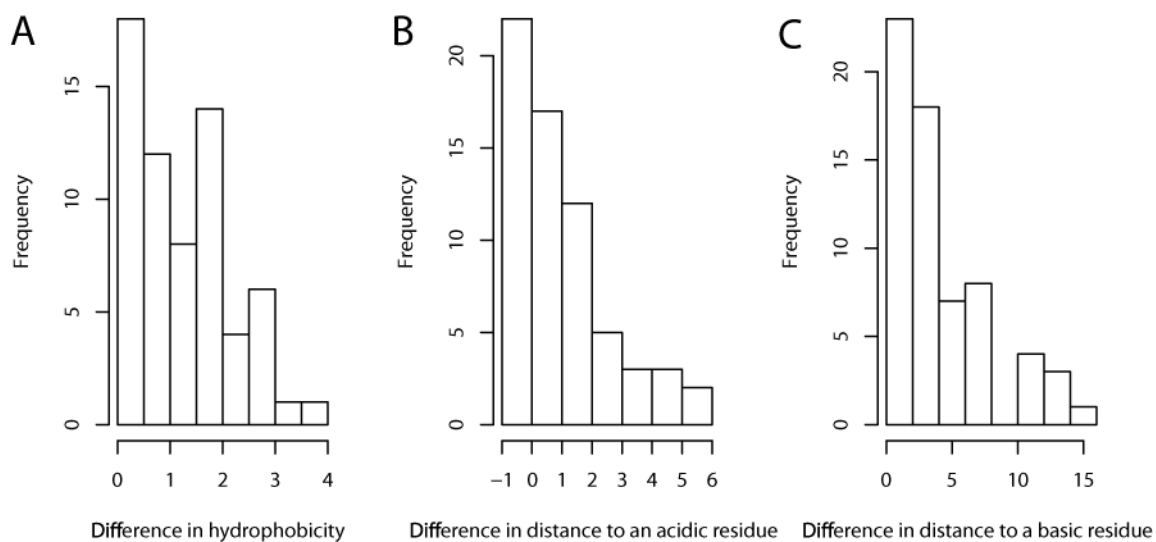


Figure A4. 1. Physico-chemical properties of phosphopeptide isomers separated by reverse phase LC.

Differences in physico-chemical properties (hydrophobicity and electrostatic forces) of 64 phosphopeptide isomers identified in IEC-6 and J774 cells. A. Difference in absolute local hydrophobicity scale between the region centered on the phosphorylated sites (windows of 3 amino acids, amino acids scale values from Kyte J., Doolittle R.F.[391]). Difference in proximity (in amino acids) of the phosphate group to an acidic residue (B) or a basic residue (C) that could affect electrostatic interactions.

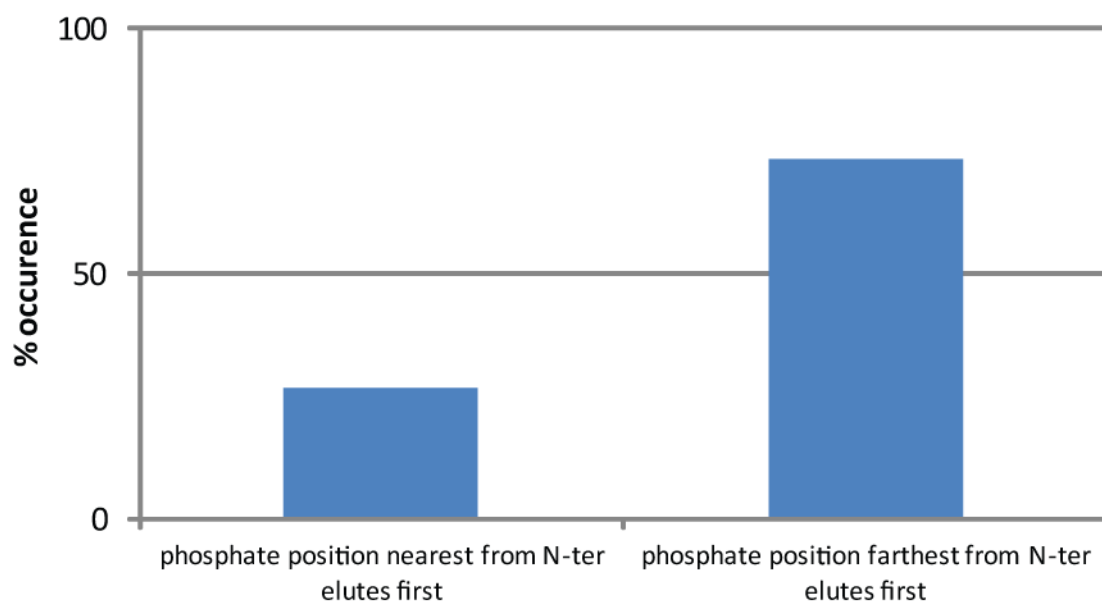


Figure A4. 2. The order of elution of phosphopeptide isomers is related to proximity of the phosphorylated residue to the N-terminus.

For 73% of isomeric pairs, phosphopeptides with the modified residue closest to the N-terminus (N-ter) elute last, suggesting that electrostatic interactions occur between the N-ter amino group and the negatively charge phosphate moiety.

I

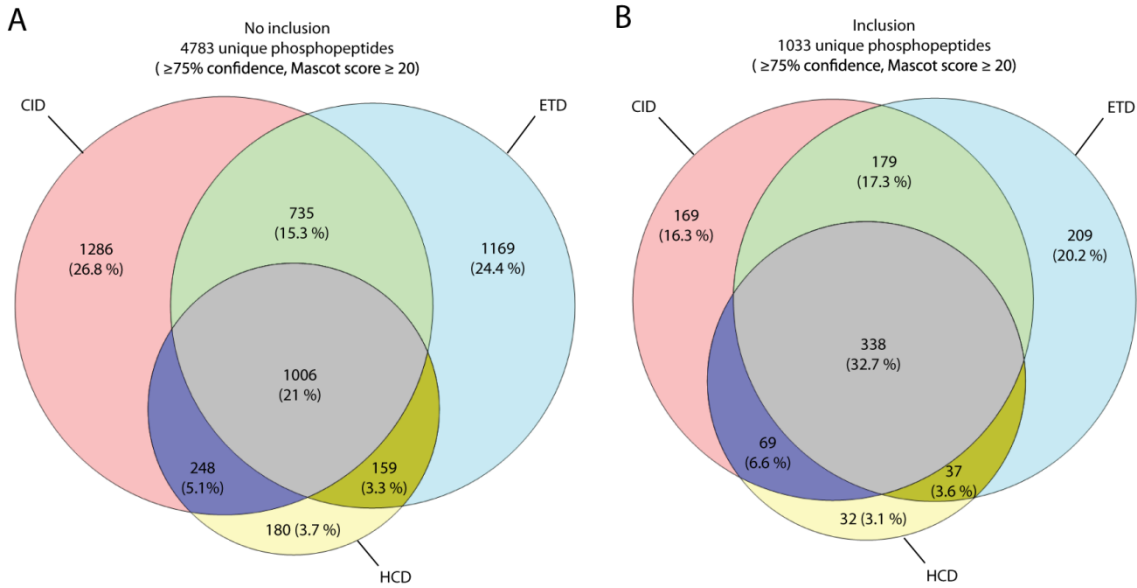


Figure A4. 3. Overlap in phosphopeptide identification between CID, HCD and ETD for *D. melanogaster* S2 phosphoproteome.

Unique peptide identification with a Mascot score > 20 and a phosphosite position confidence > 0.75 were selected. A. Number of unique peptides identified during the DDA experiment. B. Number of unique peptides identified in the targeted experiment using inclusion list and all three fragmentation modes for each precursor ion.

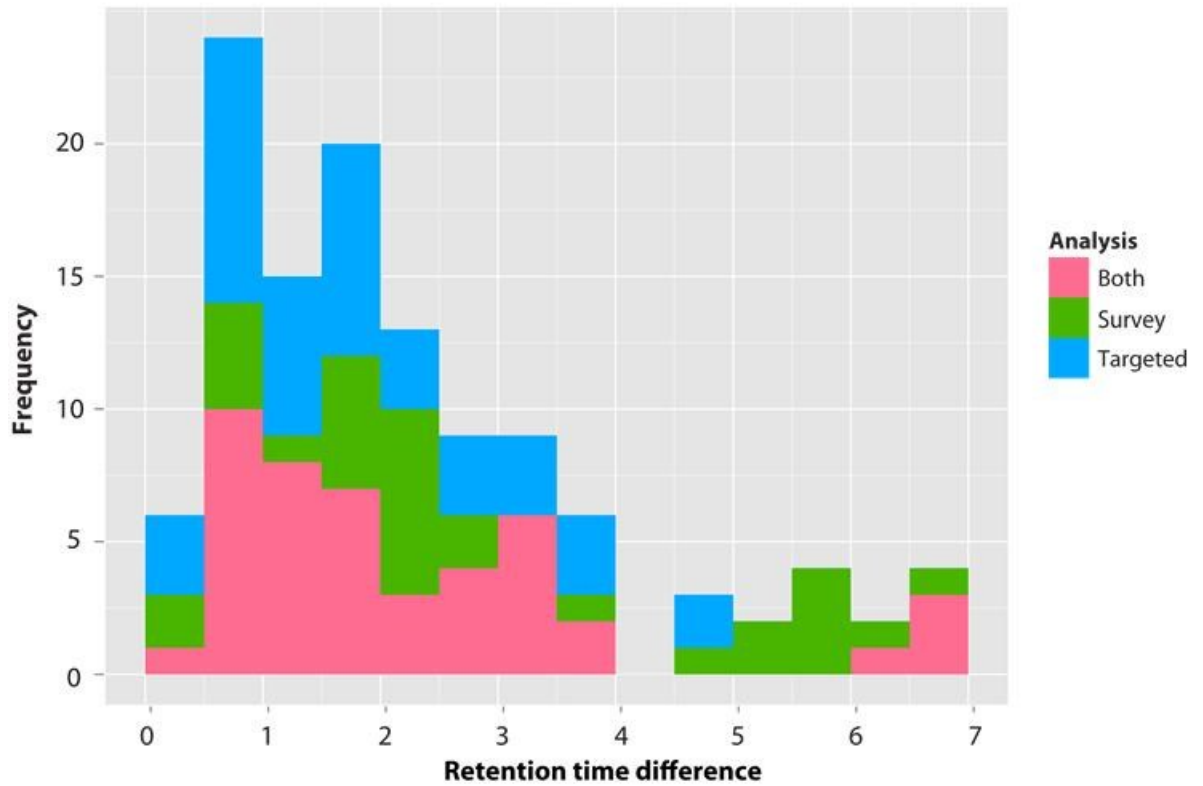


Figure A4. 4. Difference of retention time for 117 phosphopeptide isomers of *D. melanogaster* S2 cells using DDA (survey) and targeted MS/MS analyses.

Data were obtained using the LC-MS elution profile detection algorithm and assignments were confirmed by MS/MS. The histogram shows the common and uniquely detected isomers from the survey and the targeted analyses.

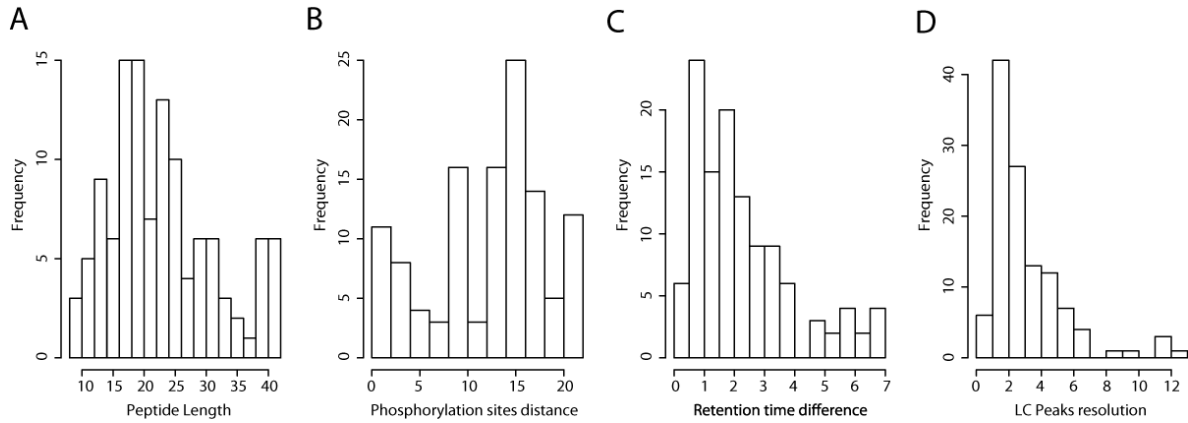


Figure A4. 5. Properties of 117 phosphopeptide isomers of *D. melanogaster* S2 cells.

Histograms for the distribution of the peptide length (A), distance in amino acids between phosphorylated site position (B), retention time difference at peak top (C), and chromatographic resolution of isomer peaks for an intensity threshold of 10 000 counts (D).

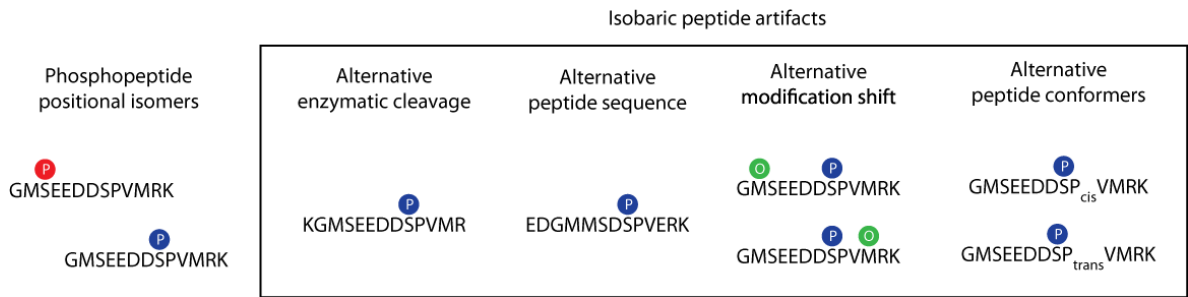


Figure A4. 6. Isobaric peptide artifacts.

Isobaric peptide artifacts can be confounded with phosphopeptide isomers. Alternative enzymatic cleavage, peptide sequence, other modification (e.g. alternate position of oxidation) and conformers are peptides with the same mass that can be separated by LC.

GIMEEIEMRP(S)LSDR (625.6³⁺)

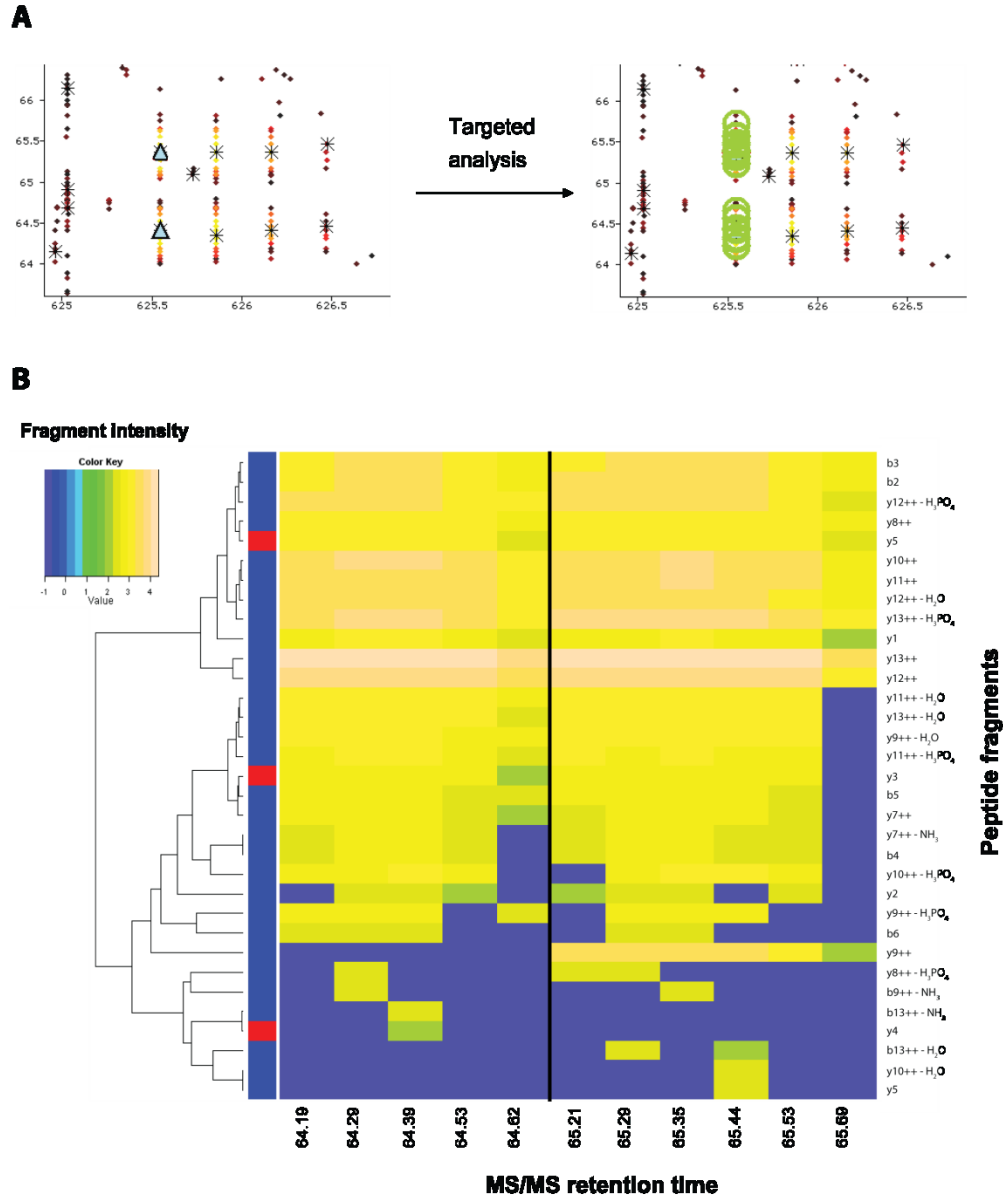


Figure A4. 7. LC separation of phosphopeptide conformers.

Example of separated phosphopeptide conformers: GIMEEIEMRp(S)LSDR (m/z : 625.6³⁺) with both methionines oxidized. A. LC-MS profile showing two separated conformers and the acquired HCD MS/MS spectra for the targeted analysis (green circle). B. Heatmap of the observed peptide fragments. Each row represents a fragment and each column a MS/MS scan. Fragments marked with red localize the phosphorylated residue. The vertical black line separates both conformers.

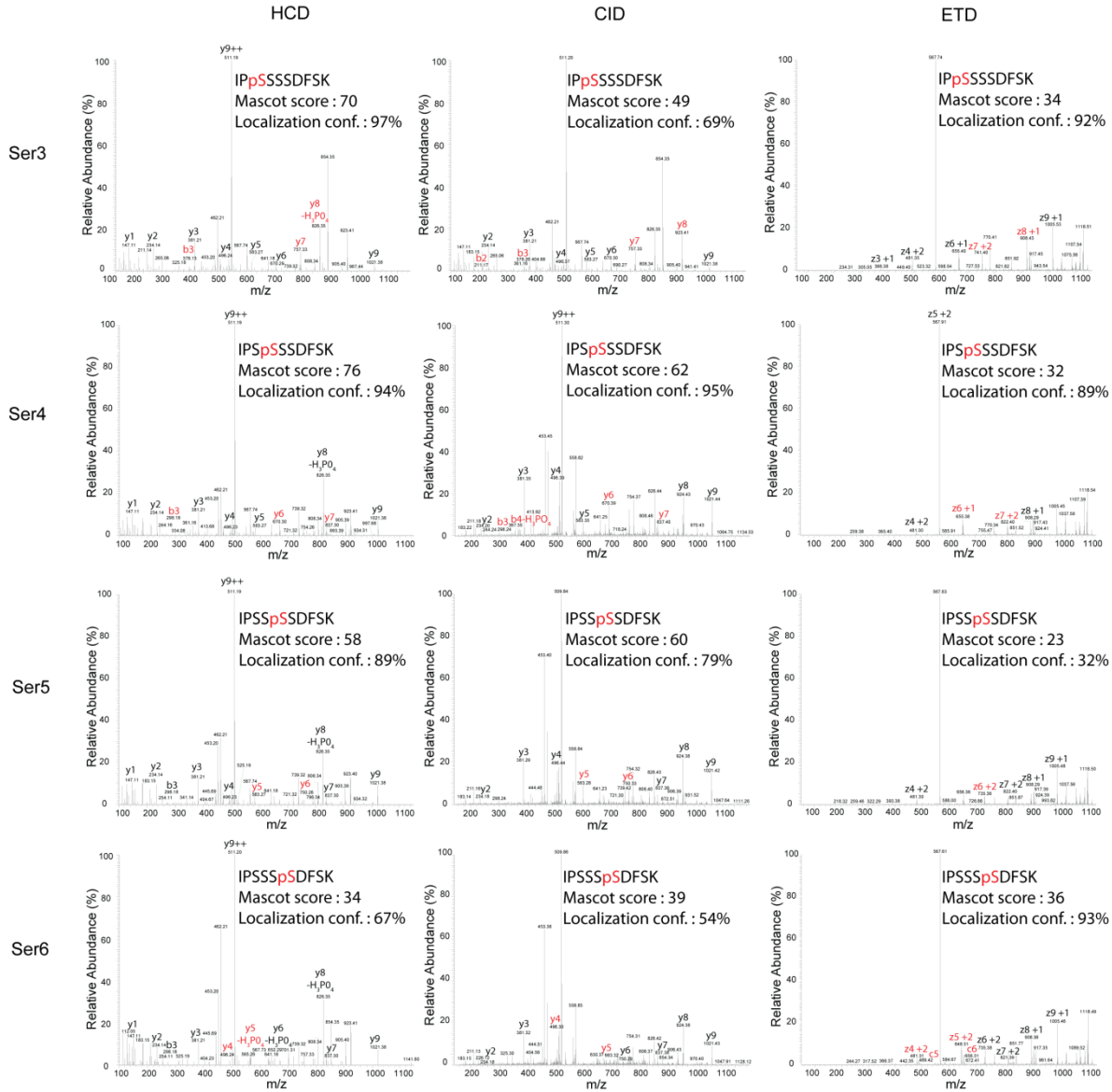


Figure A4. 8. MS/MS spectra of four phosphopeptide isomers from IPSSSDFSK.

CID, ETD and HCD MS/MS spectra are shown for the four phosphopeptide isomers of IPSSSDFSK. Mascot score and phosphorylation localization confidence are reported for each spectrum.

Table A4- I. List of all phosphosites identified in IEC-6, J774 and S2 cells. (CD-ROM).

The table reports 7 935, 3 529 and 5 596 unique phosphosites identified in IEC-6, J774 and S2 cells, respectively. Accession, gene symbol, description, species, residue, position, localization confidence, peptide sequence, peptide modifications and mascot score are reported for each phosphorylated site.

Table A4- II. Phosphopeptide positional isomers discovered in mouse and rat. (CD-ROM).

List of the 64 isomers detected from the mouse and rat datasets. Peptide sequence, modifications, m/z , localisation confidence, score, retention time, resolution of LC separation, distance between phosphorylated amino acid, distance to the nearest acid or basic residues and local hydrophobicity scale are reported.

Table A4- III. List of synthetic phosphopeptides. (CD-ROM).

List of the 9 synthetic phosphopeptides used to test the separation of isomers by RP-HPLC and the algorithm performance. Peptide sequence, modification, m/z , localisation confidence, score, retention time and resolution of LC separation are reported.

Table A4- IV. Phosphopeptide positional isomers in drosophila S2 cells. (CD-ROM).

List of the 117 isomers detected from the fly dataset and included peptide ions for the targeted analysis. In addition to the data fields reported in **Table 4-II**, inclusion m/z , inclusion time window, SCX fraction, inclusion & identification success, true isomers & ambiguous cases, and artifacts are reported for the survey and targeted analysis. MS/MS identifications from CID/ETD/HCD fragmentation for the phosphorylated peptide IPSSSSDFSK are also reported.

Annexe 5 : Contributions scientifiques

Publications

2012

- **Bridon G, Bonneil E, Muratore-Schroeder T, Caron-Lizotte O, Thibault P.** Improvement of Phosphoproteome Analyses Using FAIMS and Decision Tree Fragmentation. Application to the Insulin Signaling Pathway in *Drosophila melanogaster* S2 Cells. *J Proteome Res.* 2012 Feb 3;11(2):927-40.
- **Gaëlle Bridon, Éric Bonneil, Olivier Caron-Lizotte, Pierre Thibault.** A combined phosphoproteomics and RNAi approach to identify CK2 substrates following insulin stimulation of *drosophila* S2 cells. *Manuscrit en préparation pour J Proteome Res.*
- **Courcelles M, Bridon G, Lemieux S, Thibault P.** Occurrence and detection of phosphopeptide isomers in large-scale phosphoproteomics experiments. *J Proteome Res.* 2012. Jul 6;11(7):3753-65.

2010

- Trost M, **Bridon G**, Desjardins M, Thibault P. Subcellular phosphoproteomics. *Mass Spectrom Rev.* 2010 Nov-Dec;29(6):962-90.

Conférences et présentations

2012

- **Gaëlle Bridon**, Éric Bonneil, Olivier Caron-Lizotte and Pierre Thibault. Quantitative phosphoproteomics and RNAi experiments using FAIMS-LC-MS/MS unveil unexpected CK2 targets in insulin signaling pathway. Montreal Mass Spectrometry Discussion Group (MMSDG). *Poster presentation*. Montréal (janvier 2012) (*Prix de meilleure présentation par affiche*).

2011

- **Gaëlle Bridon**, Éric Bonneil, Olivier Caron-Lizotte and Pierre Thibault. Quantitative phosphoproteomics and RNAi experiments using FAIMS-LC-MS/MS unveil unexpected CK2 targets in insulin signaling pathway. American Society for Mass Spectrometry (ASMS) Conference. *Poster presentation*. Denver, CO, USA. (mai 2011).
- **Bridon, G.**, Bonneil, E., Caron-Lizotte, O., Thibault, P. Identification de nouvelles cibles de CK2 impliquées dans la voie de l'insuline par une nouvelle approche phosphoprotéomique combinant RNAi et FAIMS-LC-MS/MS. *Présentation orale*. Concours Simon-Pierre Noël. Département de Biochimie, Université de Montréal, Montréal (mars 2011) (*prix Merck Frosst de meilleure présentation orale*).

2010

- **Gaëlle Bridon**, Éric Bonneil, Tara Muratore-Schroeder, Olivier Caron-Lizotte, Brenda Kessler and Pierre Thibault. Comprehensive phosphoproteomics analyses of insulin signalling pathway in *Drosophila melanogaster* using high field asymmetric waveform ion mobility spectrometry and CID/ETD ion activation on a LTQ-Orbitrap. *Oral presentation*. The 24th Annual Tandem Mass Spectrometry Workshop, Lake Louise, Alberta. (décembre 2010).
- **Gaëlle Bridon**, Éric Bonneil, Tara Muratore-Schroeder, Brenda Kessler and Pierre Thibault. Enhanced Sensitivity and Coverage in Large-Scale Phosphoproteomics Studies using High Field Asymmetric Waveform Ion Mobility Spectrometry and Decision Tree Fragmentation. Troisième journée scientifique de l'IRIC. *Poster Présentation*. Montréal. (novembre 2010). (*prix de meilleure présentation par affiche*).
- **Éric Bonneil, Gaëlle Bridon**, Tara Muratore-Schroeder, Brenda Kessler and Pierre Thibault. Enhanced Sensitivity and Coverage in Large-Scale Phosphoproteomics Studies using High Field Asymmetric Waveform Ion Mobility Spectrometry and Decision Tree Fragmentation. American Society for Mass Spectrometry (ASMS) Conference. *Poster presentation*. Salt Lake City (mai 2010). (*non présente à la conférence*).
- **G. Bridon**, M. Marcantonio, M. Desjardins et P. Thibault. Dissection moléculaire des événements de signalisation précoces reliés à l'interféron gamma dans des cellules de macrophages J774 par une étude quantitative phosphoprotéomique différentielle. Retraite du Département de biochimie de l'Université de Montréal. *Présentation orale*. Saint-Adèle (mai 2010).

- **G. Bridon**, M. Marcantonio, M. Desjardins et P. Thibault. Dissection moléculaire des événements de signalisation précoces reliés à l'interféron gamma dans des cellules de macrophages J774 par une étude quantitative phosphoprotéomique différentielle. *Présentation orale*. ACFAS, Montréal (mars 2010). (*prix de meilleure présentation orale*).

2009

- **Gaëlle Bridon**, Malha Sahmi, Marc Therrien and Pierre Thibault. Molecular dissection of Ras-MAPK signaling pathway using phosphoproteomics and RNA interference. *Poster presentation*. American Society for Mass Spectrometry (ASMS) Conference. Philadelphia, PA, USA (mai 2009).
- **Gaëlle Bridon**, Malha Sahmi, Marc Therrien and Pierre Thibault. Molecular dissection of Ras-MAPK signaling pathway using phosphoproteomics and RNA interference. *Poster presentation*. Systems Biology Symposium (IRIC). Montreal (novembre 2009).

2008

- **Gaëlle Bridon**, Malha Sahmi, Marc Therrien and Pierre Thibault. Molecular dissection of Ras-MAPK signaling pathway using phosphoproteomics and RNA interference. *Poster presentation*. Ile journée de la recherche de l'IRIC. Montreal (novembre 2008).

- **Gaëlle Bridon**, Maria Marcantonio, Matthias Trost, Michel Desjardins and Pierre Thibault Phosphoproteome profiling approaches to monitor global cell signalling changes in macrophages. *Poster presentation*. Retraite du département de Biochimie de l'Université de Montréal. Saint-Adèle. (mai 2007).