Design, development and application of new technological approaches in subcellular proteomics

By

Daniel Gauthier

Division of Experimental Medicine

Faculty of Medicine

McGill University, Montreal

August 2007

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

"*The process of scientific discovery is, in effect, a continual flight from wonder.*"

Albert Einstein

"*Science is a wonderful thing if one does not have to earn one's living at it.*"

Albert Einstein

Table of contents

CHAPTER 2: Utilization of a new biotinylation reagent in the development of a nondiscriminatory investigative approach for the study of cell surface proteins

Acknowledgements

First and foremost, I would like to thank Dr. Claude Lazure for the quality of the supervision and mentorship he provided during this journey. By giving me the liberty to explore and the guidance to stay on track, he made my stay in his laboratory an instructive and fulfilling one and kept me challenged and motivated throughout my development into an independent scientist.

I would also like to thank Drs. Majambu Mbikay, André De Lean, Christian Deschepper, Puttaswamy Manjunath, David Lohnes and Gaétan Thibault for monitoring and orienting my progress during committee meetings and for sharing their experience and insights.

I would like to thank members of the laboratory, Dany Gauthier and Nadia Rabah who, along with Dr. Eric Bergeron, supported and helped me through my studies and rapidly became more friends than colleagues.

I would also like to acknowledge members of the IRCM technological platforms and services, Eric Massicotte, Martine Dupuis, Dr. Denis Faubert and his colleagues, Annie Vallée and Richard Cimon whose tremendous work and assistance were instrumental to the quality of this thesis.

Also, this work could not have been accomplished without the financial contribution of CIHR, PENCE, FRSQ, and McGill University

Finally, a special mention of appreciation must go to my dear fiancée Mélanie, my parents, my family and my friends whose collective encouragements and unfailing support in every step of the way provided me with the energy to reach my goals.

Abstract

The field of subcellular proteomics aims to describe and analyze all the proteins present in a precise subcellular compartment at a given time. In contrast to whole-cell or whole-organism proteomics, the analysis of individual organelles has provided simpler proteomes from which relevant biological information could be more easily derived. To date, the protein complement of several subcellular structures, including the mitochondria, lysosome, peroxysome, phagosome and nucleus has been described. The powerful and rapidly evolving instrumentation as well as the development of biochemical and bioinformatics tools now allow scientists to derive whole organelle models based on the proteomic data generated.

This field however still faces numerous challenges. Among those is the analysis of membrane-associated proteins, whose large and hydrophobic character complicate their extraction, subsequent separation and analysis by mass spectrometry. Another emerging limitation in subcellular proteomic resides in the difficulty in collecting enough material of a very pure preparation of the organelle of interest, which still depend on lengthy and labor-intensive density-based centrifugation as the method of choice for subcellular fractionation.

The work presented in this thesis describes the development of new methods for subcellular proteomics that address the above-mentioned limitations, and their application to relevant biological models. In chapter 2, we present the design of a non-discriminatory investigative approach to study membrane proteins. Relying on detergent-free and gel-

free procedures, this strategy allowed identification of hundreds of cell surface-exposed proteins from freshly ejaculated bovine spermatozoa, as presented in chapter 3.

Diverging from traditional cell fractionation protocols, we have refined in chapter 4 a fluorescence-assisted organelle sorting method and have employed it to acquire the proteome of corticotropes-derived secretory granules. Our procedure is not only simpler and faster, but also the amount and quality of the data we obtain compares advantageously to that generated from classical protocols.

Both methods presented in this thesis can be used separately or in combination, and can be adapted to most biological models, making them excellent new tools to perform, complete and refine subcellular proteomic work.

Résumé

Le domaine de la protéomique subcellulaire vise à décrire et analyser toutes les protéines présentes dans un compartiment subcellulaire précis à un temps donné. Contrastant avec la protéomique des cellules ou d'organismes complets, l'analyse d'organelles individuelles a généré des protéomes plus simples desquels l'information biologique pertinente peut être plus facilement extraite. À ce jour, le complément de protéines de plusieurs structures subcellulaires, incluant la mitochondrie, le lysosome, le peroxysome, le phagosome et le noyau a été décrit. L'évolution rapide et la puissance de l'instrumentation disponible couplées au développement d'outils biochimiques et bioinformatiques permettent maintenant aux scientifiques de générer des modèles d'organelles complets basés sur les données générées par la protéomique.

Ce domaine, cependant, fait toujours face à plusieurs défis. Parmi ceux-ci, on doit mentionner l'analyse des protéines associées à la membrane dont la taille et l'hydrophobicité compliquent l'extraction, la séparation subséquente et l'analyse par spectrométrie de masse. Une autre limitation émergeante en protéomique subcellulaire est l'obtention d'une préparation très pure d'organelles d'intérêt et ce, en quantité suffisante, qui dépend toujours de longues et laborieuses centrifugations basées sur la densité comme méthode de choix pour la fractionnement subcellulaire.

Le travail présenté dans cette thèse décrit le développement de nouvelles méthodes en protéomique subcellulaire qui s'adressent aux défis mentionnés précédemment, et leur application à des modèles biologiques pertinents. Dans le chapitre 2, nous présentons l'élaboration et la mise au point d'une approche investigatrice non

discriminatoire pour étudier les protéines de membrane. Basée essentiellement sur des procédures sans détergent et sans gel de séparation, cette stratégie a permis l'identification de centaines de protéines exposées à la surface de spermatozoïdes bovins fraîchement éjaculés, tel que présenté au chapitre 3.

S'éloignant des protocoles traditionnels de fractionnement cellulaire, nous avons élaboré au chapitre 4 une méthode de tri d'organelles basée sur la fluorescence permettant l'utilisation d'un trieur automatisé et nous l'avons employée pour définir le protéome des granules de sécrétion des cellules corticotropes. Notre procédure n'est pas seulement plus simple et plus rapide, mais la quantité et la qualité d'information que nous obtenons se comparent très avantageusement à celles générées par les protocoles classiques.

Chacune des méthodologies présentées dans cette thèse peut être utilisée individuellement ou en combinaison, et peut être adaptée à la plupart des modèles biologiques, ce qui en fait d'excellents nouveaux outils pour produire, compléter et raffiner du travail de protéomique subcellulaire.

Contribution of authors

CHAPTER 2: Utilization of a new biotinylation reagent in the development of a non-discriminatory investigative approach for the study of cell surface proteins. (*Proteomics*, 2004 Dec;4(12):3783-90)

Bernard F. Gibbs performed the LC-MS/MS analysis of the samples, Nadia Rabah was responsible for the Sf9 cell culture, Claude Lazure provided funding and mentorship for the project. Daniel J. Gauthier was responsible for all other aspects of the project, including sulfo-NHS-iminobiotin synthesis, labeling experiments, LC-MS/MS sample preparation and MS results analysis. It must also be mentioned that a patent with the names of the two inventors DJ Gauthier and C Lazure was deposited but, following examination, it was ruled that, due to a technicality, it was not receivable.

CHAPTER 3: Determination of the cell surface proteome of freshly ejaculated bovine spermatozoa using a non-discriminatory investigative approach. (Draft manuscript from preliminary results, in preparation)

Dany Gauthier performed labeling and peptide purification experiments, Claude Lazure provided funding and mentorship for the project. Daniel J. Gauthier was responsible for all other aspects of the project, including characterization of PC4 by 1D and 2D gel electrophoresis, labeling and peptide purification experiments and LC-MS/MS data analysis.

CHAPTER 4: Organellar proteomics in a day: Flow cytometry-assisted purification and proteomic analysis of the corticotropes dense-core secretory granules. (Manuscript submitted to *Proteomics*).

Jacqueline A. Sobota generated the AtT-20-PHM-mGFP cell line and performed immunocytochemistry experiments, Francesco Ferraro did preliminary FACS experiments, Richard E. Mains supervised JA. Sobota and F. Ferraro and agreed to provide us with the AtT-20-PHM-mGFP cells, Claude Lazure provided funding and mentorship for the project. Daniel J. Gauthier was responsible for all other aspects of the project, including cell culture, immunocytochemistry, subcellular fractionation, optimization of the fluorescence-assisted organelle sorting, generation of LC-MS/MS samples and LC-MS/MS data analysis.

List of figures and tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of abbreviations

1D: one-dimensional

2D: two-dimensional

ABPP: affinity-based protein profiling

ACTH: adrenocorticotrophic hormone or adrenocorticotropin

ADAM: a desintegrin and metalloprotease

BAC: benzyldimethyl-n-hexadecylammonium chloride

BLAST: basic local alignment search tool

BSP : bovine seminal plasma

CCV: clathrin-coated vesicle

CgA: chromogranin A

CgB: chromogranin B

CHAPS: 3-[(3-cholamidopropyl)-dimethylammonio]-1- propanesulfonate

CIAQ: Centre d'insémination artificelle du Québec

CpE: carboxypeptidase E

CRH : corticotropin releasing hormone

DIGE: differential gel electrophoresis

dsRNA: double stranded RNA

DTT: dithiothreitol

ECL: enhanced chemiluminescence

EGFP: enhanced green fluorescent protein

ER: endoplasmic reticulum

ERGIC: endoplasmic reticulum to Golgi apparatus intermediate compartment

ERp : endoplasmic reticulum protein

ESI: electrospray ionization

FACS: fluorescence-assisted cell sorting

FAOS: fluorescence-assisted organelle sorting

FSH: follicle-stimulating hormone

FT: Fourier transformed

GFP : *Aequorea victoria* green fluorescent protein

GH: growth hormone

GHRH: growth hormone releasing hormone

GnRH: gonadotropin releasing hormone

GPI : glycosylphosphatidylinositol

HBFP: human body fluid proteome

HPLC: high pressure liquid chromatography

HRP: horseradish peroxidase

ICAT: isotope-coded affinity tag

ICR: ion cyclotron resonance

IEF: isoelectric focussing

IGF: insulin-like growth factor

IMAC : immobilized metal affinity chromatography

iTRAQ: isotope tags for relative and absolute quantitation

LC: liquid chromatography

LH: luteinizing hormone

LTQ: linear quadrupole ion trap

MALDI: matrix-assisted laser desorption ionization

mGFP: *Aequorea victoria* monomeric green fluorescent protein

mL: milliliter

mRNA: messenger RNA

MS : mass spectrometry

MS/MS : tandem mass spectrometry

MudPIT: multidimensional protein identification technology

NCBI: national center for biotechnology information

NE: nuclear envelope

NET: nuclear envelope transmembrane protein

NHS: N-hydroxysulfosuccinimide

NK: natural killer

nL: nanoliter

NMR: nuclear magnetic resonance

PACAP: pituitary adenylate cyclase activating polypeptide

PAGE: polyacrylamide gel electrophoresis

PAM: peptidyl-glycine alpha-amidating mono-oxygenase

PBS: phosphate-buffered saline

PC: Prohormone convertase

PC1/3 : prohormone convertase 1/3

PC2 : prohormone convertase 2

PC4 : prohormone convertase 4

PFF: peptide fragmentation fingerprinting

PHM: peptidylglycine alpha-hydroxylating monooxygenase

pI: isolectric point

PIST: PDZ domain protein interacting specifically with TC10

PMD: piecemeal degranulation

PMF: peptide mass fingerprinting

POMC: pro-opiomelanocortin

PRL: prolactin

PTM: post-translational modification

Q: quadrupole

Rab: ras-related in brain

RNAi: RNA interference

RP: reverse phase

rRNA: ribosomal RNA

SCX: strong cation exchange

SDS: sodium dodecyl sulfate

Sf: *Spodoptera frugiperda*

Sg: secretogranin

SG: secretory granule

SILAC: stable isotope labeling by amino acids substitution in cell culture

SNARE: soluble N-ethylmaleimide-sensitive factor (NSF) attachment receptor

SOFA: single organelle fluorescence analysis

SRIF: somatotropin release-inhibiting factor

SSC: side scatter

TGN: *trans*-Golgi network

TOF: time of flight

TRH : thyrotropin releasing hormone

TSH: thyrotropin

VAMP: vesicle-associated membrane protein

VIP: vasoactive intestinal peptide

ZP: zona pellucida

β -LPH: β-lipotrophin

Chapter 1

Introduction and literature review

1.1 PROTEOMICS

1.1.1 Preface

Since its early description in 1995, the field of proteomics has undergone a tremendously fast evolution. Initially perceived as the logical and necessary sequel to the various genomics projects, proteomics was proposed to describe living cells and organisms not by their fundamental DNA blueprint, but rather by its direct output, the proteins. Although the technologies and definitions are getting increasingly refined, the goal in proteomics has remained the same: to describe as accurately and completely as possible the full complement of all proteins expressed in a specific cell, tissue, organ or organism under defined conditions.

Comprising over 50% of the dry weight of a typical cell, proteins are major players involved at one point or another in every biological process. Ranging from a few to a few thousands amino acids in length, these products from mRNA translation exhibit a wide range of physicochemical properties that allow them to perform optimally their functions. Generally, proteins are classified into families and subfamilies according to the structural similarities of the domains they share. Moreover, these proteins, by virtue of their interaction with other molecules within the cell, have often evolved to perform a specific function and act, for example, as catalysts, signal receptors, switches, structural components, motors or pumps[1]. Therefore, to fully understand a living cell or organism, it is not enough to know its genes. It is also crucial to be able to identify which of these genes are ultimately translated into functional protein effectors, where these proteins are located, and in what proportions. Figure 1.1 illustrates the "Omics" cascade comprising complex databases that, as a group, aim to comprehensively describe the response of a

biological system to disease, genetic and environmental perturbations[2]. Systems biology is a field which aims to integrate these various sources of information to study the interactions between components of biological systems and how these interactions give rise to the function and/or behaviour of that system[3].

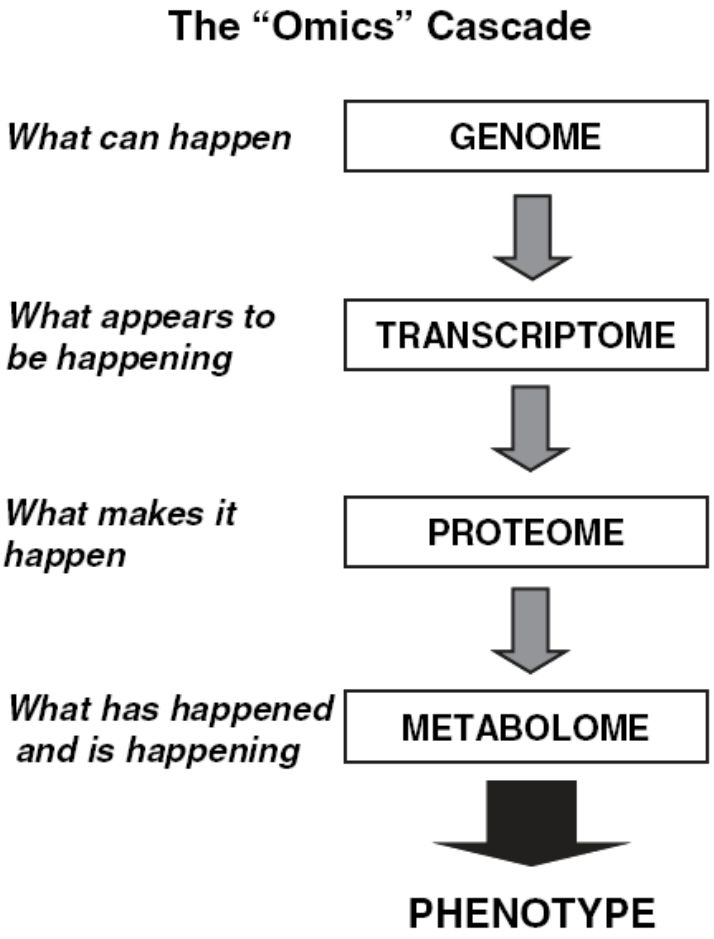


Figure 1.1: The "Omics" cascade[2]. The sequential and integrated response of organisms to their environment involves alterations at the levels of DNA, mRNA, proteins and metabolites. Complex databases are each informative at their specific level on these changes that culminate in an alteration of the phenotype.

1.1.2 Challenges in proteomics

Over the last 12 years, the field of proteomics has been faced with many challenges, the most serious of which being related to the nature and complexity of the samples to be analyzed. Even if native proteins are biopolymers of essentially the same twenty natural amino acids, the exact combination of these residues can generate an almost infinite range (20^X where x is equal to the residue number) of polypeptides exhibiting enormous variation in their physicochemical properties. Separation and analysis of those proteins is therefore equally complex. A detailed description of the tools and methods currently employed in proteomics is presented in section 1.2. However, it is first noteworthy to discuss a few of the specific challenges that need to be addressed in the field.

1.1.2.1 Sample preparation

The biological relevance of the data generated in proteomics is most often linked to the initial quality of the protein sample to be analyzed. Therefore, sample preparation is the critical prerequisite step in determining the quality of the proteomic data collected. Firstly, it is imperative that the biological sample be as enriched as possible for the model of interest, be it a whole organism, a cell or an organelle. Then, methods must be employed to maximize the number of individual proteins as well as the crude amount of proteins collected from the sample, while minimizing as much as possible the other contaminant molecules (lipids, carbohydrates, salts, ions, etc.) that would interfere with the analysis. And although it might sound obvious, sample preparation still remains one of the biggest challenges in the field, be it for collection or for unambiguous validation of the data.

1.1.2.2 Protein sample complexity

As mentioned above, the intrinsic properties of the proteins in a biological sample can be quite varied.

1.1.2.2.1 Size and shape

Proteins are large linear biopolymers of amino acids generated from the translation of messenger RNA by ribosomes. They come in sizes and shapes that can be as varied as their individual amino acid sequence. Ranging in size from a few to over twenty-six thousand amino acids for a mass of nearly three million daltons[4], they also fold into stable tertiary structures by virtue of electrostatic, van der Waals, hydrogen bonding and hydrophobic forces.

1.1.2.2.2 Isoelectric point

Also determined by its primary sequence, the isoelectric point (pI) of a protein represents the pH at which its global charge, i.e. the sum of the individual charges of all its constituent amino acids side chains and its amino and carboxy termini will be zero. It is a property of a protein that is evolutionary derived and is usually related to its function and subcellular localization[5]. In this case again, pI of proteins in a single sample can range from 2 to 13, a fact that will need to be considered when analyzing proteomes.

1.1.2.2.3 Post-translational modifications, degradation

Most proteins bear one or many covalent post-translational modifications (PTMs) that will affect their behaviour in a cell. PTMs are generally performed by specialized

enzymes and are essential in modulating numerous cellular processes, including DNA transcription, signalling, motility, and protein degradation to name a few. As they affect a protein's charge, molecular weight and chemical behaviour, PTMs such as phosphorylation, glycosylation, sulfation, oxidation and acetylation add to the complexity of a sample, especially since the same protein can now take multiple forms. In the same vein, proteolysis needs to be considered. Where regulated proteolysis can be essential to a protein's function, proteolytic degradation of the sample, on the other hand, must be prevented.

1.1.2.2.4 Solubility, hydrophobicity, hydrophilicity

A protein's hydrophobic or hydrophilic character will be a major determinant of it's solubility in various solvents. Hydrophilic proteins are usually easily extractable in detergent-free buffers and amenable to further proteomic analysis. On the other hand, proteins having large hydrophobic segments, such as multiple-pass transmembrane proteins will require detergents and are prone to aggregation and precipitation. Even if proteomic analysis of this subset is still very challenging, increasing our repertoire of tools for the study of membrane proteins is essential, as they are absolutely crucial to cell function and are most important targets as far as drug design is concerned[6].

1.1.2.2.5 Dynamic range of expression

A single serum/plasma sample can contain over one thousand proteins. But more importantly, they can vary in abundance by a factor as high as 10^{10} copies[7]. So, methods must be developed to deplete the most abundant proteins such as albumin,

immunoglobulins and fibrinogen in order to expose the less abundant, but usually more interesting candidates.

1.1.2.3 Data collection, interpretation and validation

Proteomic analysis generates enormous amounts of computerized data. A single sample run in a large scale proteomics analysis can provide information on over 10 000 peptide queries, themselves yielding hundreds of thousands of MS/MS fragmentation spectra. Bioinformatic tools and search engines such as Mascot™[8] need to use this data to probe existing protein databases in order to achieve protein identification. The quality of these tools, algorithms and databases is of paramount importance. Once identification is established, the experimenter must go back to the biological model to validate the information using conventional biochemical methods while keeping efficiency in mind, since it is neither realistic nor desirable to perform thousands of western blots or knock-outs to prove the validity of proteomic approaches.

1.1.3 Bottom-up and top-down proteomics

The term “bottom-up” or “shotgun” proteomics refers to the group of methods, techniques and tools that are employed in order to identify the proteins present in a biological sample from a mixture of their proteolytic peptides. It usually involves i) an extraction step, in which proteins are released from the sample, ii) a purification and separation step, in which the proteins are isolated from the rest of the molecules in the sample, iii) a proteolytic digestion step, in which peptides are generated from whole proteins and iv) an identification step which combines peptides separation by chromatography, peptides ionization, and mass analysis of these proteolytic fragments.

Using bioinformatics tools, the sequence of the final peptides can be deduced and the identity of the proteins in the initial sample can be determined, thereby using its small peptidic fragments (bottom) to reconstruct a complete proteome (up). Bottom-up proteomics is the strategy of choice in most studies, even if the final mixture of peptides is extremely complex since recent improvements in both separation and mass analysis tools can usually cope with this problem.

Another strategy, called "top-down proteomics", rather deals with intact proteins in the sample, generally reducing the need for purification, proteolysis and separation steps, thereby minimizing sample loss. However, the chromatographic separation, ionization and mass measurements of large intact proteins still pose technical challenges that will need to be resolved for this approach to be more commonly utilized[9]. Among the most serious problems faced are related to sensitivity and throughput[9]. Whereas known protein standards can be analyzed with ultrahigh sensitivity with fast acquisition of fragmentation data[10], dealing with wild type unkown proteins bearing variable number of PTMs is proving more difficult. Sample complexity is a strong limiting factor in top-down proteomics, and the use of MALDI for ionization both hinders throughput and sequencing abilities.

1.1.4 Smaller proteomes for simplicity

The description of all the proteins contained in a sample is the first crucial step in orienting proteomics-based studies toward the elucidation of biological questions. The early proteomics work (~6000 publications (1985-2001)) was aimed at cataloging proteins and building databases, leading to an understanding of the intrinsic complexity in controlling mechanisms of cellular function in higher organisms[11]. Nowadays, not only is

this important work being pursued, but it is also being refined. In order to simplify the data and to ask and answer more targeted questions, the elucidation of smaller, simpler or more targeted proteomes represent an interesting and feasible endeavour.

The first way to achieve this is by selecting a small biological sample. For example, analysis of the human body fluid proteome (HBFP) has as its main objective the deciphering of the proteome of individual body fluids in order to identify potential biomarkers important in health and disease[12]. This is perceived as a powerful approach that could find clinical applications rapidly as far as diagnosis and prognosis is concerned. In the search for biomarkers of cardiovascular disease, quantitative mass spectrometry (MS) methods have allowed he limiting of the number of candidates to 177 proteins out of over 1000 proteins in the human plasma or serum[7]. Targeted proteomics analysis of the human serum also led to the identification of putative biomarkers for over 12 types of cancer[12]. The analysis of urine exosomes has permitted the identification of 295 proteins, many of which being the product of genes know to be associated with renal diseases[13]. Other fluids, including cerebrospinal, saliva and tear fluids were also subjected to analysis and the recent results were reviewed[12].

Also, only specific sections of an organ or tissue can be analyzed. However, extreme care must be taken to avoid contamination by surrounding cells if an accurate proteome of the desired target is to be generated. One of the most efficient ways to accomplish this is by using laser-assisted microdissection techniques, which provide the advantage of allowing selection of the target cells and sections based on their morphologies as well as generating high quality protein extracts[14]. Using this method, distinct matrix-assisted laser desorption ionization (MALDI) mass spectrometry spectra from normal and malignant epithelia breast cells were generated[15], with reproducible

spectra obtained from as little as 10 cells. Also, using two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) and staining, 17 out of the 675 proteins detected showed tumor-specific alterations between normal squamous cells and squamous cell carcinoma cells acquired by microdissection[16].

Finally, when possible, another way to limit biological sample complexity is to deal with a single cell type. In human, it usually translates into components of the blood/serum. For example, descriptive studies on the erythrocyte proteome have now identified over 500 proteins present in this cell type[17]. Also, a comparative proteomic analysis of normal and sickle-cell anaemia red blood cells using 2D-differential gel electrophoresis (DIGE) revealed that 44 protein spots originating from 22 unique protein sequences presented at least a 2.5 fold difference in abundance between the two models[18]. It suggested a role for proteins involved in actin binding, free radical control and protein repair and turnover in the disease. Another model, the platelets, has received a lot of attention due to its physiological role in blot clotting, as well as its pathological involvement in thrombosis and heart disease. Within the last 10 years, the number of distinct protein isoforms identified in platelets has gone from 25 to over 2000 following proteomic analysis[19, 20]. This vast improvement is mostly explained by the technological advances in gel-based protein separation and especially MS-assisted protein identification, illustrating the importance of possessing powerful and accurate tools in proteomics.

Other relatively simple biological samples amenable to analysis are pathogens. Numerous strains of yeast, bacteria, virus and even parasites have been the subject of proteomics studies, either to obtain a comprehensive proteome or to identify more specific pathogenic factors. Strain-specific markers are also of great interest, as rapid

discriminatory identification of these organisms could contribute to diagnosis at the clinic level. Among interesting examples are the work performed on the opportunistic *Candida albicans*, that causes more and more infections especially in immunosuppressed patients[21]. Also studies on various strains of bacteria, including *Haemophilus influenzae*[22], *Helicobacter pylori*[23, 24] and *Mycobacterium tuberculosis*[25] are leading to increasingly complete proteome databases. Among the most interesting applications of these studies is the discovery of strain and strain-specific host biomarkers as well as the elucidation of the proteins important for host-pathogen interactions, leading to potential vaccine and therapeutic targets[26]. Even if proteomic study of parasitic agents is still in its early days, this field is getting rapidly organized and the term "parasitoproteomics" has been defined as the study of the reaction of the host and parasite genomes through the expression of the host and parasite proteomes during their complex biochemical cross-talk[27]. The sum of the recent proteomic work on *Plasmodium falciparum*, the parasite responsible for most cases of human malarial infections and death, has now identified over half of the 5300 proteins predicted from its genome and monitored both quantitive and qualitative changes throughout the parasite's life cycle[28]. It also shed light on the response of the pathogen to drug treatment[29]. Finally, the application of proteomics to the study of viruses has already covered virion composition, viral protein structures, virus-virus and virus-host protein interactions, and changes in the cellular proteome upon viral infection, as recently reviewed[30].

1.1.5 Functional proteomics

Another approach to reduce the complexity of the sample is to focus on a specific subset of proteins exhibiting distinct properties, most often related to their role in the cell.

The subfield of functional proteomics is particularly focussed upon such groups of polypeptides.

1.1.5.1 Activity-based protein profiling

Enzymes are proteins that catalyze specific chemical reactions in the cell. Depending on the nature of the reaction, they are classified into families including proteases, hydrolases and phosphatases to name a few. The activity of these enzymes is regulated at various levels. Activity-based protein profiling (ABPP) is aimed at developing chemical probes to label the active site of actives enzyme in order to perform functional proteomics[31]. Following treatment of total proteomes with the site-specific tagged chemical probes, only enzymes from the desired class or family whose active site is available (ie, active and not inhibited) will be labelled. Retrieval of the enzymes of interest is accomplished by affinity purification using the tag on the probe, resulting in a much restricted proteome that can be more easily analyzed. Recently, a high-throughput method combining ABPP and multidimensional protein identification technologies (MudPIT) has permitted the identification of over 50 enzyme activities in breast cancer[32]. Enzyme substrates can also be studied in an analogous manner. Following pharmaceutical inhibition of the activity of specific enzymes, proteomics can be used to identify their endogenous targets. Using this quantitative pharmacoproteomic approach, several substrates of matrix metalloproteinases in a breast cancer cell line were identified[33].

1.1.5.2 Glycoproteomics

Glycosylation is a PTM that involves addition of carbohydrate groups to specifically targeted asparagine (N-linked) and serine, theonine and tyrosine (O-linked)

residues of a protein. This modification is present in all eukaryotes, most prevalent at the cell surface and plays various roles in protein folding, cell structure, cell-cell adhesion and signalling[34]. Consequently, isolating and analysing the subset of glycosylated proteins represent an endeavour of interest. However, the complexities of the carbohydrate structures as well as the diversity of the linkage of this modification complicate the selective analysis of these proteins. Among the techniques employed to tackle this task, affinity binding to lectin allows selective retrieval of glycopeptides from a proteolytic digest. MS can then be used to identify proteins and glycosylation sites. Such method has allowed identification of 400 unique glycosylation sites from 250 proteins in a *Caenorhabditis elegans* protein extract[35]. A large scale human plasma glycoproteomic study combining hydrazide chemistry-based retention of the glycosylated proteins with deglycosylation and LC-MS/MS resulted in the identification of over 2000 peptides from 303 glycoproteins[36].

1.1.5.3 Phosphoproteomics

Phosphorylation is another crucial PTM implicated in regulating nearly all cellular processes, including transcription, transport, signalling and cytoskeletal rearrangements. Dynamic addition and removal of phosphate groups at selected serine, threonine and to a lesser extent tyrosine residues is performed by a plethora of kinases and phosphatases acting in an integrated and regulated manner to modulate cellular response to stimuli. The subfield of phosphoproteomics utilizes various methods in an attempt to focus on phosphorylated proteins, which could represent between one third and one half of all proteins at a given time[37]. If the model studied is amenable to radiolabeling, incorporation of ^{32}P or ^{33}P will expose this subset of proteins upon autoradiography. Anti-

phosphoserine, anti-phosphothreonine and anti-phosphotyrosine antibodies are nowadays increasingly available for immunoblot detection. Finally, at the gel level, fluorescence-based methods such as the Pro-Q Diamond[38] stain can be employed to specifically reveal phosphorylated proteins. Moreover, immobilized metal affinity chromatography (IMAC) can be employed to retrieve phosphoproteins and phosphopeptides and lead, following chemical derivatization and MS/MS analysis, to their identification[39].

1.1.5.4 Other post-translational modifications

Oxidation of amino acids, by reactive oxygen species or reactive nitrogen species is another PTM that is implicated in signal transduction processes. The most recent information on the possible roles and proteomic analysis of these less frequent modifications has been reviewed[40]. Ubiquitination, the addition of polymers of a conserved 76 amino acids peptide to the side chain of lysines in targeted proteins by ubiquitin ligases plays a central role in protein degradation, but also in stabilization, localization, interactions as well functional activity of many protein substrates[41]. Strategies for large scale proteomic analysis of proteins and peptides bearing this modification rely on a signature peptide motif originating from tryptic digestion and comprising a glycine-glycine fragment bound to the protein lysine residues by an isopeptide bond that can be detected by MS. These signature peptides can be modified by fluorescent affinity tags for further enrichment. This approach has allowed identification of 126 peptides targeted to the proteasome in cells deficient in the Rpn10 polyubiquitin receptor[42]. Recently, an improved, more comprehensive strategy for MS-based identification of ubiquitination sites was developed. Using three different tags on both the peptides' internal lysine residues and C-termini has led to the the identification of 96

ubiquitination sites in proteins purified from human MCF-7 breast cancer cells, a marked improvement in the number of sites that could be identified using standard approaches[43].

1.1.6 The lower end of proteomics: peptidomics and metabolomics

The field of peptidomics can be viewed as bridging the gap between proteomics, aimed at a description of the complement of proteins in a system and metabolomics, aimed at the analysis of all metabolic products and by-products generated by cellular activity. The peptidome therefore represents the complement of all small proteins and peptides generated by the activity of various proteases in a biological sample. Since it is effectively a subset of protein products resulting from biological activity, it could also, in the context of this section, be considered as a restricted functional proteome. It is certainly of great biological relevance since restricted proteolytic processing is a way to modulate the activity of numerous peptide hormone precursors. For example, the proopiomelanocortin (POMC) precursor can be processed differently to yield distinct peptide hormones depending on the cell type in which it is produced[44] (also, Figure 1.12). All the above-mentioned polypeptides are quite small and their analysis is better addressed by peptidomics. The most sought-after peptidome is certainly that of human blood, again in the quest for relevant and sensitive biomarkers. A review of the most recent tools and results in blood peptidomics has been published[45]. Interestingly, neuroendocrinology is also starting to benefit from peptidomic studies. Recently, a combination of liquid chromatography (LC), MS and bioinformatics has described the peptidome of the pancreatic islets of Langerhans, thereby revealing novel peptides acting in the regulation of blood glucose levels[46]. The quest for the peptides responsible for the obesity phenotype in carboxypeptidase E (CpE) deficient mice was also conducted using

neuropeptidomic strategies[47]. The human is not the only model studied, and a similar approach has permitted the identification of 28 neuropeptides, including eight brand-new, from *Drosophila melanogaster* larvaes[48].

1.1.7 Conclusion

As described above, the field of proteomics, albeit relatively young, is quickly expanding to enormous proportions. Among the many contributing factors, one of the crucial elements, as we will see in the forthcoming section, certainly resides in the power and accuracy of the analytical tools, especially related to MS. Indeed, these no longer represent a bottleneck and so scientists are exploring multiple aspects of proteomics, including tools and biological models, to further advance our understanding of living systems.

1.2 TOOLS IN SUBCELLULAR PROTEOMICS

Over the last 12 years, many tools and methods have been developed in order to overcome the challenges involved in identifying the full complement of proteins of a living system. Tremendous progress in the technologies related to MS-based protein and peptide identification has been accomplished, alleviating this specific limitation and allowing the analysis of large and complex samples. As illustrated in section 1.3, the most groundbreaking subcellular proteomic studies were not necessarily the ones identifying the most proteins, but also the ones presenting innovative approaches to prepare the sample or analyze the collected data. In this section, the tools and methods most widely employed in proteomics will be reviewed, with particular focus on subcellular proteomics.

1.2.1 Sample preparation tools

In subcellular proteomics, the quality of the initial protein sample is a key element of a successful study. The sensitivity and resolution provided by MS, although instrumental in the identification of interesting candidates, will also identify the contaminating proteins present in the sample. Consequently, the goal in such studies is to analyze a sample originating from a preparation that is as enriched as possible in the organelle of interest, while containing minimal amounts of the various contaminants described in section 1.2.1.3.2 below.

1.2.1.1 Cell lysis

Cell lysis is the initial step in subcellular proteomics. Cell disruption can be performed by any of the four following methods: mechanical, pressure, ultrasound and freeze-thaw-based homogenizations. Mechanical homogenization is by far the most widely employed method in subcellular proteomics and uses rotators-stators as tools. This method is especially versatile since it can be employed for a variety of tissues and cells, and the use of disposable tips completely eliminates cross-contamination between sample sets[49]. These tips can be blunt and present a fixed and selectable clearance between them and the tube, as in Potter homogenization, or can simply be cutting tools, allowing control in the stringency of the homogenization. Sonicators on the other hand generate high energy ultrasound waves that will destroy cells upon impact, while pressure homogenizers will force cells in suspension to pass through a small orifice under high pressure, resulting in cell lysis. However, both these methods usually indiscriminately destroy cells and subcellular structures, and are mostly employed to obtain whole-cell proteomes for more resistant systems such as microbial and plant cells[50]. Finally, freeze-

thaw homogenization takes advantage of the formation of ice crystals in water solutions to lyse the cells. It is usually less stringent and is perceived as a complementary method for those mentioned above to ensure maximal protein recovery.

1.2.1.2 Subcellular fractionation

Cells are extremely complex systems, and even though the human genome is estimated to encompass between 25000 and 30000 genes[51], a single cell could contain more than 100000 proteins species when including cleavage products, splice variants and PTMs[52]. Since actual analytical tools can not efficiently cope with this level of complexity, ways of simplifying proteomes are essential. One of the most efficient methods to achieve this is through subcellular fractionation, which leads not only to simplification of the proteome, but also to an enrichment of low-abundance proteins by 3-8 times[53]. The latter comes about through the collection of enriched fractions amenable to further analysis allowing more starting material of interest to be used. Subcellular fractionation, developed over five decades ago[54], has well proven its efficiency in biochemistry and cell biology. It comprises essentially two steps. The first one is cell disruption, by one of the above-mentioned methods. It is then followed by fractionation of the homogenate into distinct populations of organelles and structures. Most of the time, cell lysis is followed by brief centrifugation at low speeds to remove nuclei, unbroken cells and large debris.

1.2.1.2.1 Density-based fractionation

In nearly all subcellular proteomic studies, as examplified in section 1.3.1, the fractionation method that is almost universally employed involves several rounds of density-based centrifugation. This is based on the fact that the nature and composition of the various subcellular compartments confer them a specific density that is relatively distinct from that of other organelles. Density centrifugation can be subdivided in two types. In velocity centrifugation, the cell lysate is applied to a relatively shallow gradient and the components will migrate at different speed toward the bottom of the tube under centrifugal forces according to their size and shape. After a predetermined amount of time, the separated components can be recovered by piercing the bottom of the tube. In equilibrium centrifugation, most widely employed in subcellular proteomics, the lysate is applied to a steeper gradient and the particles will move up or down the gradient until their density matches that of their surroundings[1]. Once separation is complete, the fraction of interest is collected and can be directly analyzed or submitted to a new round of centrifugation on a gradient of different composition for further separation.

The nature of the gradient will affect the centrifugal behaviour of the subcellular compartments. Some of the required characteristics of centrifugation media take into account the facts that they should not alter the cells or organelles while providing a useful dynamic density range using reasonable concentrations. Toxicity and osmolarity changes, on the other hand, must be avoided as much as possible[55]. Sucrose is still the media of choice for building gradients[52], although some of its properties, including high osmolarity and viscosity, can represent a problem in certain circumstances[55]. Consequently, alternative compounds have been developed to circumvent these issues. A synthetic polymer of sucrose known as Ficoll® was introduced to minimize hyperosmotic effects,

but is not entirely successful at every concentrations unless compensated by a salt gradient[56]. Dextran, a natural sucrose polymer, has been used to fractionate cell but it is very viscous[57]. Many other molecules such as iodinated compounds (Metrizamide®, Nycodenz®, Optiprep®) and colloidal silica (Percoll) are also employed, each with its own characteristics but all aiming at achieving maximal separation while minimizing convection forces and sample alterations[58]. It is worth mentioning that gradients can be continuous, in which case density increases linearly along the tube, or discontinuous in which case the gradient is divided into portions of consecutive fixed densities. Among the advantages of density-based centrifugation for cell sample fractionation, we note that it can achieve nearly complete separation of particles in a complex mixture according to their size or density using a determined combination of duration, centrifugal force and density range of the gradient. However, this method of fractionation presents important caveats including the variable success encountered in enrichment and/or depletion of the sought after organelle as well as the co-isolation of cytoskeletal components with the organelles. Because of these reasons, and since minimal contamination can now be detected, density-based centrifugation, although still the method of choice, is increasingly perceived by experts in the field as one of the major limitations in subcellular proteomics[59, 60].

1.2.1.2.2 Affinity-based fractionation

Unfortunately, very few alternative and efficient fractionation methods that would be compatible with proteomics are available. One of them is certainly the affinity isolation of specific subcellular compartments. As its name suggests, this method is based on the affinity between a molecule present on the organelle and a molecule located on a

solid support such as beads or columns. As an example, cell membranes have been labelled through biotinylation enabling one to take advantage of the strong affinity of biotin for avidin. Doing so to recover the labelled membranes from a lung cancer cell line using streptavidin-coated magnetic beads, it was possible to identify of 898 proteins, nearly 85% of them annotated as being associated with the membrane[61]. Although this approach was efficient in this context, we can hardly envision this type of specific labelling in a whole cell lysate. Immunoaffinity is therefore a preferred method for fractionation of organelles. Briefly, it consists of incubating the cell lysate with an antibody directed at a protein specifically exposed at the surface of the subcellular structure of interest. Then, the antibody-labeled lysate is exposed to a secondary antibody bound to magnetic beads. Upon exposure to a magnetic field the ensuing complexes will be separated[62]. Used alone or in combination with density-based fractionation, this procedure has provided satisfactory results in the proteomic study of peroxysomes[63], plasma membranes[64] and synaptic vesicles[65]. However, despite the fact that it generated highly enriched preparations, affinity fractionation is still only marginally employed as it also presents numerous disadvantages. Among those, we have to mention the cost, the relatively low bead binding capacity refractory to large scale preparation and last, but not least, is the requirement for detergents to recover the organelles from the solid support due to the extremely high affinity of binding. In turn, it will bring into the sample the antibodies and will require prior separation of the proteins by 1D- or 2D-gel electrophoresis rather than direct digestion and LC-MS/MS. However, there is a growing tendency to favor gel-free protocols in the field for large scale studies as it is much more rapid and efficient.

1.2.1.2.3 Model-specific strategies

A few compartment-specific fractionation strategies exist and will be discussed herein. However, even if they have proven their efficiency for their specific subcellular compartment, they can unfortunately not be adapted to others, rendering them of quite limited usefulness. The first spectacular demonstration is the latex-bead incorporation-based purification of the phagosomes[66, 67]. The very low density of the phagosomes containing a latex bead allow them to surface in a density gradient, allowing purification in a single step instead of successive recovery of selected fractions embedded into gradients. Further discussed in section 1.3.1.6, this rapid and efficient method led to the elaboration of a thorough proteome within approximately 5 years. Another example of model-specific strategy, presented in section 1.3.1.9, concerns the purification of plasma membranes by treatment of cell lysates with digitonin. This reagent changes the membranous components' buoyant density and improves separation of the plasma membrane from other membranes[68], as confirmed by the activity of various membrane-associated enzymes[69]. Both of these examples belong to the very few modifications to the standard protocols that have been proposed to change the way subcellular fractionation is conducted.

1.2.1.2.4 Fluorescence assited organelle sorting

Fluorescence assisted cell sorting (FACS) is routinely employed to perform selective enrichment of cell populations bearing specific markers based on fluorescence[70, 71]. In 1985, Murphy's group first described the concept of single organelle fluorescence analysis (SOFA) and subsequently suggested its application to sort single organelles[72]. The term fluorescence assisted organelle sorting (FAOS) has also been introduced to

describe the use of flow cytometry to sort subcellular structures and components. However, this represents a challenging task as the organelles to be sorted are routinely 10 to 100 times smaller than the original cells. Furthermore, the resistance of the target organelle to the shear and tear fluid forces in the cell sorter is limited. Nonetheless, publications have described sorting of organelles such as mitochondria[73] and phagosomes[74] using various fluorescent probes. However, due to technical constraints, most of these were analytical in nature though some preparative studies still achieved a significant level of enrichment[75]. Endosomes were of particular interest since they can be accessed from outside the cell and loaded transiently with fluorescent membrane dyes or fluorescently labeled ligands under different conditions[76]. This organelle was therefore studied very early[77] and also more recently[78] by combining standard fractionation protocols with flow cytometry-assisted organelle sorting. Interestingly however, very little work on FAOS was published since the mid 1990s. Moreover, no large scale proteomic study was published to date with this approach. This is surprising, since the technique has shown its potential for organelle purification although the technical constraints at the time clearly prevented the obtention of a pure preparation of intact structures. The level of sophistication in modern flow cytometers[71] and the dire need for new subcellular fractionation techniques have prompted us to explore this alternative with great success, as described in chapter 4.

As a final word on subcellular fractionation, it must be mentioned that assessment of the purity of the final organelle preparation must always be thoroughly performed. In order to achieve this, researchers usually rely on a battery of classical biochemical techniques as well as on electronic microscopy. On the biochemical aspect, the two most employed tools for fraction analysis are immunoblotting and enzymatic assays. Both

methods obviously rely on the association of known markers with specific subcellular compartments. Immunoblotting is quite rapid and sensitive, but artefactual redistribution of some proteins due to cell disruption can complicate interpretation of the results. The presence of a specific compartment is also routinely evaluated by measuring the activity of resident enzymes such as β-1,4-galactosyltransferase (Golgi[79]) or β-hexoseaminidase (lysosome[80]) in the different gradient fractions. Both of these methods allow quantification of the relative enrichment or depletion ratio for the organelle of interest, since separation is never complete. Finally, electron microscopy allows direct visual observation of the composition of the fraction from which purity if the sample can also be assessed.

1.2.1.3 Protein sample preparation

Once the organelles of interest have been isolated, the next step is to obtain their proteins in conditions that are compatible with further separation an analytical steps.

1.2.1.3.1 Solubilization of proteins

Protein solubilization is one of the most crucial steps in subcellular proteomics and will be a major determinant of whether an experiment will lead to the successful acquisition of a complete proteome. Despite the great diversity in the proteins' physicochemical properties mentioned earlier (see section 1.1.2.2) and the presence of various contaminating molecules in biological samples, the goal is to achieve simultaneous solubilization of all the proteins. Many strategies have addressed the problem of solubilization of "unattainable" proteins such as membrane[81, 82], very low and very high molecular weight[83, 84] and extreme pI proteins[85] for proteomic analysis.

However, one must keep in mind that this process is sample-dependent and never 100% efficient, so a combination of solubilization methods is sometimes required.

The three classes of agents employed for solubilization are chaotropes, detergents and reductants but they are not all compatible with all subsequent analytical steps. Chaotropes disrupt hydrogen bonds and hydrophilic interactions enabling proteins to unfold with all ionizable groups exposed to solution[49]. Urea is the chaotrope of choice in proteomics as it is compatible with isoelectric focusing (IEF) and 2D-electrophoresis, and it can be further combined with thiourea to maximize protein solubilization[86]. As far as detergents are concerned, there are several available. These amphipathic hydrophobic interaction disruptors can be classified as ionic (ie, SDS), non-ionic (ie Triton-X100) and zwitterionic (ie CHAPS). The nonionic or uncharged detergents are the most popular these days in proteomics, being stable over a large pH range, strong enough to effectively solubilize and prevent aggregation of proteins, while not interfering with IEF or gel electrophoresis. On the other hand, SDS is readily applicable if the proteins are subsequently resolved by conventional 1D-SDS-PAGE. Finally, the role of reductants is to maximize protein unfolding by cleaving disulfide bonds. Initially, sulfhydryl reducing agents such as dithiothreitol (DTT) were employed, but they were shown be charged and migrate out of pH gradients during IEF. DTT was recently replaced by uncharged phosphines which do not interfere with further alkylation steps. Reductants and other solubilizing agents have been previously reviewed[87]. It is at this point essential to mention that most of the above-mentioned procedures for protein solubilization are applicable if they are followed by gel-based separation methods. For gel-free proteomics, care must be taken not to introduce substances in the sample that would interfere with liquid separation unless prior removal strategies are available.

1.2.1.3.2 Removal of contaminants

Contaminating molecules present in the protein extract can interfere with or completely prevent further analytical steps such as IEF, LC or MS. It is therefore important to remove those contaminants while minimizing protein loss. These contaminants can be of various origins and include detergents, salts, lipids, nucleic acids and polysaccharides. Although numerous contaminant-specific removal strategies have been developed over the years, recent availability of commercial kits has improved large scale sample preparation. As far as detergents are concerned, dialysis is efficient but discriminatory protein loss is a concern, especially in low-concentration samples. SDS removal kits based on detergent precipitation (SDSAways™ (Protea Bioscience), SDS-Out SDS precipitation reagent™ (Pierce)) are now increasingly utilized. Salts are usually removed by dialysis, protein precipitation[88] or commercial kits[89] (ReadyPrep 2-D clean-up kit (Biorad)). Also, if 2D-gel electrophoresis separation is to be performed, the use of centrifugal filter devices and of CHAPS in the sample buffer allows for efficient removal of lipids and salts, thus ensuring high percentage of protein recovery and high-quality separation. Alternate protocols for delipidation involving, for example, a reversed phase (RP) C_{18} solid phase extraction cartridge were also tested successfully[90]. Polysaccharides and nucleic acids tend to cause proteins streaking on 2D-gels leading to poor results[91]. The former are usually removed by protein precipitation while treatment of the sample with protease-free RNAses and DNAses is efficient in removing large polymers of nucleic acids[92].

1.2.1.3.3 Protein depletion and enrichment

In most cases, the need for enrichment and depletion of low and high abundance proteins respectively is reduced or eliminated by the isolation of single organelles in subcellular proteomics, which already constitutes an important enrichment process. However, selection of subsets of post-translationally modified proteins in a subcellular compartment can be achieved with the same strategies as those described in section 1.1.5. Such a selection process was used and led to the recent elucidation of the mitochondrial phosphoproteome revealing 84 phosphorylation sites on 64 proteins[93]. Once more, it must be mentioned that despite the availability of methods of preparation, there is no single strategy that can be universally employed on all samples. Therefore, the experimenter must carefully select which ones will lead to a significant improvement of the final results. On the other hand, each purification step will unavoidably results in specific and non-specific sample loss, potentially fading or altering the final picture of the proteome.

1.2.2 Sample analysis tools

Once the sample of interest has been obtained, it is submitted to a series of techniques and methods aiming at identifying as accurately and completely as possible all of its constituent proteins. Figure 1.2 schematically illustrates the most frequently used steps leading to the acquisition of a proteome.

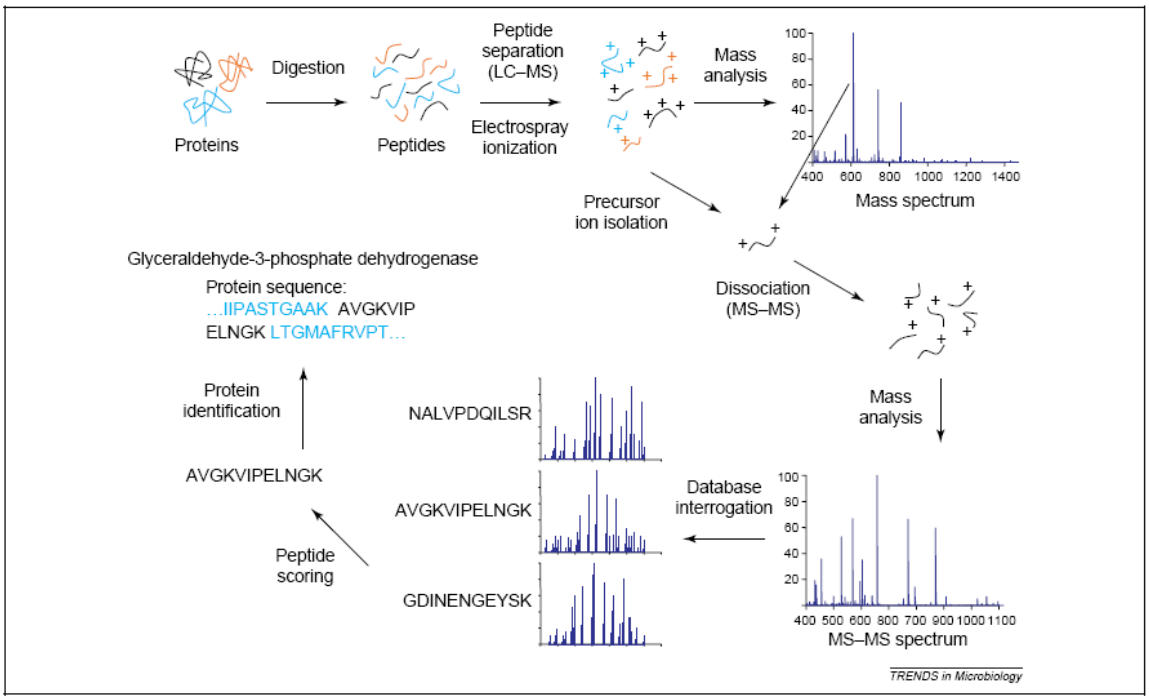


Figure 1.2: Diagram of a generalized high-throughput MS/MS proteomics pipeline[94]. Following proteolytic digestion of the proteins, the peptides are separated by LC-MS. The masses of the precursor ions and its dissociation fragments are obtained and the ensuing data is used to identify the proteins using bioinformatics tools and searching established databases.

1.2.2.1 Electrophoresis and staining

For several decades, 1D-SDS-PAGE has been and still is the most widely used method for the separation of protein samples. Developed in the 1970s, it involves separation according to their mass of a mixture of proteins submitted to an electric field through a gel matrix, resulting in a pattern of distinct bands[95]. Although it is suitable for most applications, the level of resolution that can be reached is insufficient to discriminate thousands of proteins in a single gel lane. As a consequence, two-dimensional (2D) gel electrophoresis has been developed and revolves around the principle of visualizing on the same gel proteins that have been separated according to two different physicochemical properties. In nearly all subcellullar proteomics studies using 2D-gel electrophoresis, proteins are initially separated by IEF, which separates the proteins according to their pI. In IEF, a potential is applied and the proteins migrate in a pH gradient until their net charge becomes zero, at which point they stop. The resulting separation has been greatly facilitated and stabilized by the introduction of commercially available strips wherein the pH gradients are immobilized, hence allowing separation of thousands of proteins[96]. Following isoelectric focusing, the proteins are then separated by SDS-PAGE according to their molecular weights. Upon staining, the resulting pattern is a distribution of spots, each corresponding to individual species of proteins to which an approximate pI and molecular weight can be attributed, as illustrated in Figure 1.3. The pattern is further complicated by the fact that many proteins will resolve as a series of spots since they bear several PTMs that will alter their charge or molecular weight. Following separation, the proteins can be revealed by various classical (Coomassie brilliant blue[97], silver staining[98, 99]) or the most recent fluorescent stains which overcome many of the drawbacks of the traditional stains[100]. In addition to the fact that they are

more sensitive, fluorescent dyes have allowed the development of 2D-difference gel electrophoresis (2D-DIGE)[101]. In this technique, samples to be compared are stained with two fluorescent dyes of different colors, and analysis of the overlapping spots with imaging softwares allows detection of as little as 0.5 fmol of proteins and differences of +/- 15%[101]. An exhaustive review of the different protein stains and their applications is available[102]. Numerous softwares also exist for protein spot detection and quantification and can be directly integrated in the workflow of proteomic platforms[103, 104].

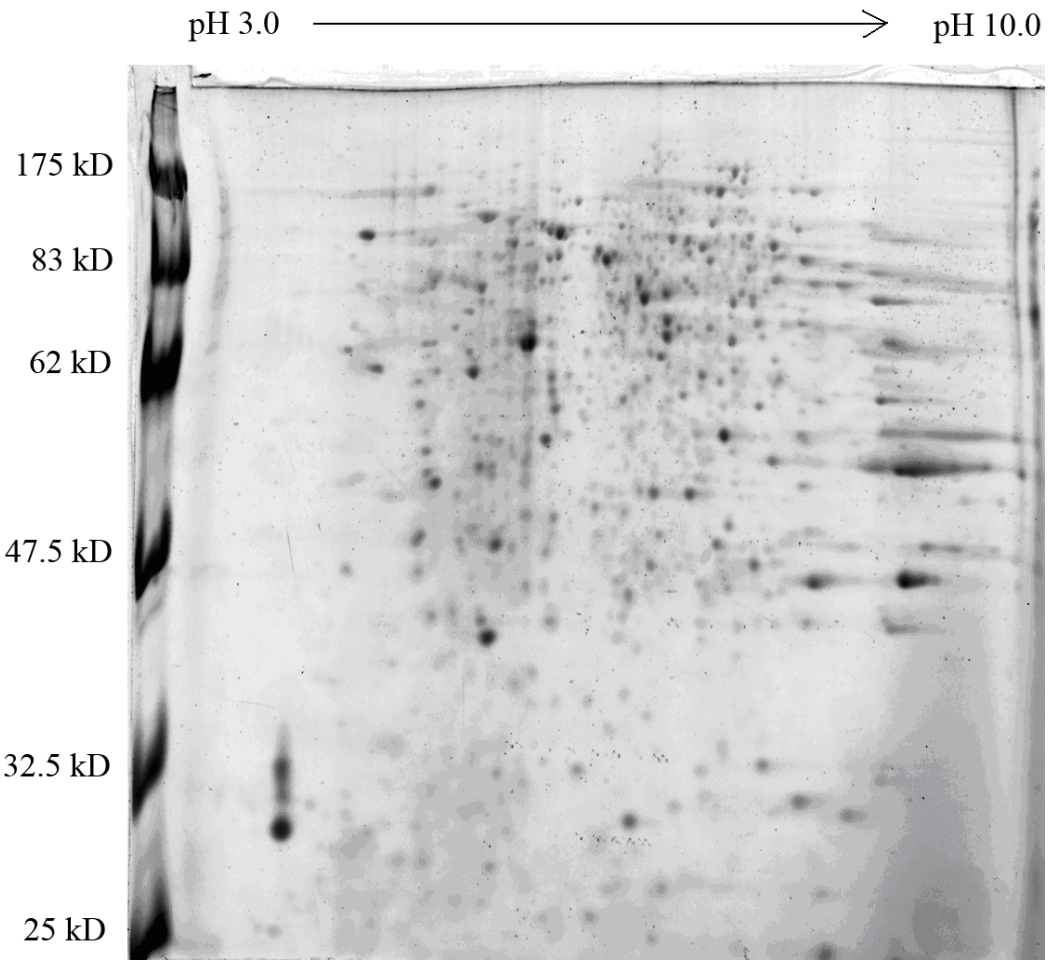


Figure 1.3: Representative two-dimensional gel electrophoresis of mouse testes protein extract. 50ug of proteins were separated on a pH 3-10 immobilized gradient strip in the first dimension and then applied to a 10% SDS polyacrylamide gel in the second dimension. Spots were reevaled by Coomassie brilliant blue staining (Gauthier and Lazure, unpublished results)

1.2.2.2 Liquid chromatography-based separation

Once the proteins have been separated by 2D-gel electrophoresis, the spots are excised from the gel and the proteins are digested into their constituent peptides by the action of endoproteases, usually trypsin. In large scale bottom-up proteomic analyses, the gel separation step can be bypassed and the total cell lysate can be directly proteolytically digested, leading to a sample of tremendous complexity comprising as many as a few million peptides[105]. Even in subcellular proteomics, where only a subset of all cellular proteins are investigated, the complexity of the peptide sample from a given spot or fraction is usually too high for direct analysis by MS. Overwhelming numbers of peptides competing for protons in the ionization step will cause a suppression effect, resulting in the non-detection of many low abundance peptides, even if the sensitivity limit of the instrument has not been reached[105]. So, in nearly all cases, the next step in most proteomic analysis is the liquid chromatographic separation of the peptides prior to their characterization by MS. The high pressure liquid chromatography (HPLC) instrumentation and procedures in proteomics are conceptually very similar to the traditional ones, the main difference residing in the flow rate, which is now reduced to the nl/min range[106]. In one dimensional (1D) liquid chromatography (1D-LC, or simply LC), reversed phase liquid chromatography is almost exclusively employed. In that case, the peptides interact with alkyl groups (C_4, C_8 or C_{18}) bound to the column matrix and their separation depends on their partition coefficient between the solid and mobile (an acidic gradient of an organic solvent) phase. This method has the advantage of offering high peak capacity while the eluate is fully MS compatible[107]. However, in some cases like shotgun proteomics dealing with hundreds of thousands of peptides, this level of separation is still not sufficient and so the combination of many separation mechanisms

(multidimensional chromatography) is required. The most widely employed combination begins with a strong cation exchange (SCX) column, separating the peptides by their positive charge, followed by a reversed phase (RP) column as shown in Figure 1.4. Trap columns or precolumns are employed to concentrate the sample and remove interfering contaminants such as salts[108]. The initial reports of multidimentional protein identification technology (MudPIT) relying on multiple orthogonal liquid separations was attributed to Yates III and coworkers[109, 110]. A detailed review of the recent advances in this powerful technique that is increasingly used to obtain thorough subcellular proteomes is available[111].

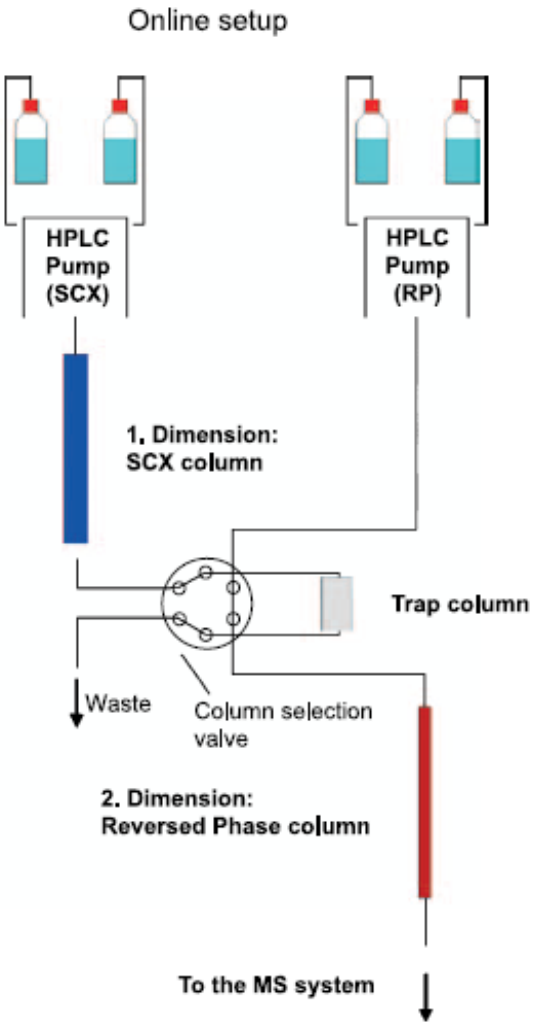


Figure 1.4: Typical configuration for two-dimensional liquid chromatography[105]. In this online system, an automated column selection valve is used to switch from the strong cation exchange to the reversed-phase columns allowing continuous flow of sample to the mass spectrometer.

1.2.2.3 Mass spectrometry

Even if the first mass spectrometer was developed over 100 years ago, the amount of sample, the sophistication of the instruments and the expertise of the scientists that were required have delayed its application as a routine method in protein chemistry until the 1980s[105, 112]. Until that time, protein identification was mainly performed by N-terminal sequencing[113]. But development of soft ionization methods (matrix assisted laser desorption ionization (MALDI)[114], electrospray ionisation (ESI)[115]) as well as improvements in the accuracy and sensitivity of mass analyzers led to an increased utilization of MS. In fact, mass spectrometers are probably the most crucial instruments in proteomics as far as protein identification is concerned. They measure the mass/charge (m/z) ratio of proteins and peptides. The information produced by the mass spectrometer, lists of peak intensities and (m/z) values, can be manipulated and compared with lists generated from ‘‘theoretical’’ digestion (“*in silico*”) of a protein or ‘‘theoretical’’ fragmentation of a peptide present in databases using various bioinformatics tools[116] as previously illustrated in Figure 1.2. Mass spectrometry comprises two consecutive steps, ionization and mass analysis.

1.2.2.3.1 Ionization

The process of ionization refers to the methods used to volatilize and charge the biomolecule. In proteomics, this is accomplished mainly by MALDI or ESI. In MALDI, the sample is dispersed and cocrystallized onto a sample plate with an organic matrix usually composed of a small organic molecule possessing an aromatic ring which can absorb energy at the wavelength of the laser. Ionization occurs by pulsed laser radiation primarily absorbed by the matrix, causing desorption and ionization of the analyte[105]. In

ESI, the liquid sample circulates through a fine needle at the tip of which a strong electric field is applied. As a result, a fine spray of small charged droplets is produced and is projected toward the mass analyzer. In both cases, charging is usually induced by addition or loss of protons to form the $M+H^+$, $M+2H^{2+}$, etc. Whereas MALDI mostly generate singly charged peptides, ESI can commonly generates multiply charged peptides. Although ionization is sufficiently efficient to lead to protein identification, both methods described have their limitations. In MALDI, not all endoproteolytic peptides will cocrystallize efficiently with the matrix and will be of a size within the detection range. In ESI, ionization will be different from one peptide to another, and therefore voltage parameters of the instrument must be adjusted for the best compromise. Moreover, too many peptide ions will lead to competition and signal suppression.

1.2.2.3.2 Mass analysis

Once ionized peptides have been generated, they are directed to the mass analyzer. In dealing with large and complex mixtures of peptide such as those generated in subcellular proteomics, obtaining the mass of all the peptides is simply not sufficient to delineate the initial protein composition of the sample. In most cases, tandem mass spectrometry (MS/MS or MS^n) is employed to obtain additional information on the peptide ions. In MS/MS, the mass of the intact peptide (precursor or mother ion) is first determined, and then this peptide is segregated from all other precursor peptides and is dissociated into smaller fragments, known as daughter ions. This is mostly accomplished via collision-induced dissociation, during which the accelerated precursor ion collides with molecules of an inert gas such as Ar or N_2[117]. A spectrum of the (m/z) ratios of the various fragments is thus obtained and all pertinent information is integrated and

compared to theoretical spectra to obtain protein identification from databases. There are four major types of mass analyzers, which can function under vacuum on their own or in combination.

Time of flight (TOF) analyzers rely on the principle that peptide ions are accelerated in an electric field and will then travel a certain distance in a field-free region, the duration of this travel being related to their m/z ratio. The ion mass can then simply be calculated by the time it takes to traverse the tube of the analyzer. Quadrupole analyzers consist of four parallel rods between which the ions circulate in a controlled and oscillating electric field which will only allow ions of precise m/z to go through without being diffracted on the rods. Scanning through the electric field allows acquisition of a mass spectrum for all ions. Fourier transformed ion cyclotron resonance (FT-ICR) MS uses high magnetic fields to trap the ions and cyclotron resonance to detect and excite the ions, with very high resolution and mass accuracy[116]. The ions are injected into a trap in line with the orientation of the magnetic field. The ions are moving in cyclotron motion and pass the detector plate inducing electrons to flow. To get a MS spectrum, the signal intensity as a function of time is Fourier transformed[105].

Finally ion trap analyzers work around the same principle as quadrupole, but they literally trap ions within a small region using an oscillating field between two cap and one magnetic ring electrode. Ions are resonantly activated and ejected by electronic manipulation of the field[118]. Ion traps are exceptionally sensitive since they can accumulate ions over a certain period of time. A new mass analyzer, the orbitrap, has been recently commercialized and presents extremely interesting characteristics[119]. As its name suggests, it is an ion trap but it operates with a completely different mechanism, without a magnet. Ions are rather trapped in an electrostatic field, in equilibrium between

the electrostatic attraction to the central tubular electrode and the centrifugational force arising from their initial velocity, causing them to oscillate in complex spiral patterns, as shown in Figure 1.5[119]. Their mass is analyzed using a FT of their oscillation frequencies. The orbitrap provides exceptional accuracy in mass measurements, usually less than 2 ppm in complex peptide mixtures[120], and great resolution, up to 100,000. The linear quadrupole ion trap (LTQ) Orbitrap was employed in our proteomic analysis of the mature dense-core secretory granules (SGs) in chapter 4. As mentioned above, mass analyzers are usually used in combination in order to precisely analyze both precursors and fragmentation ions and even post-translational modifications.

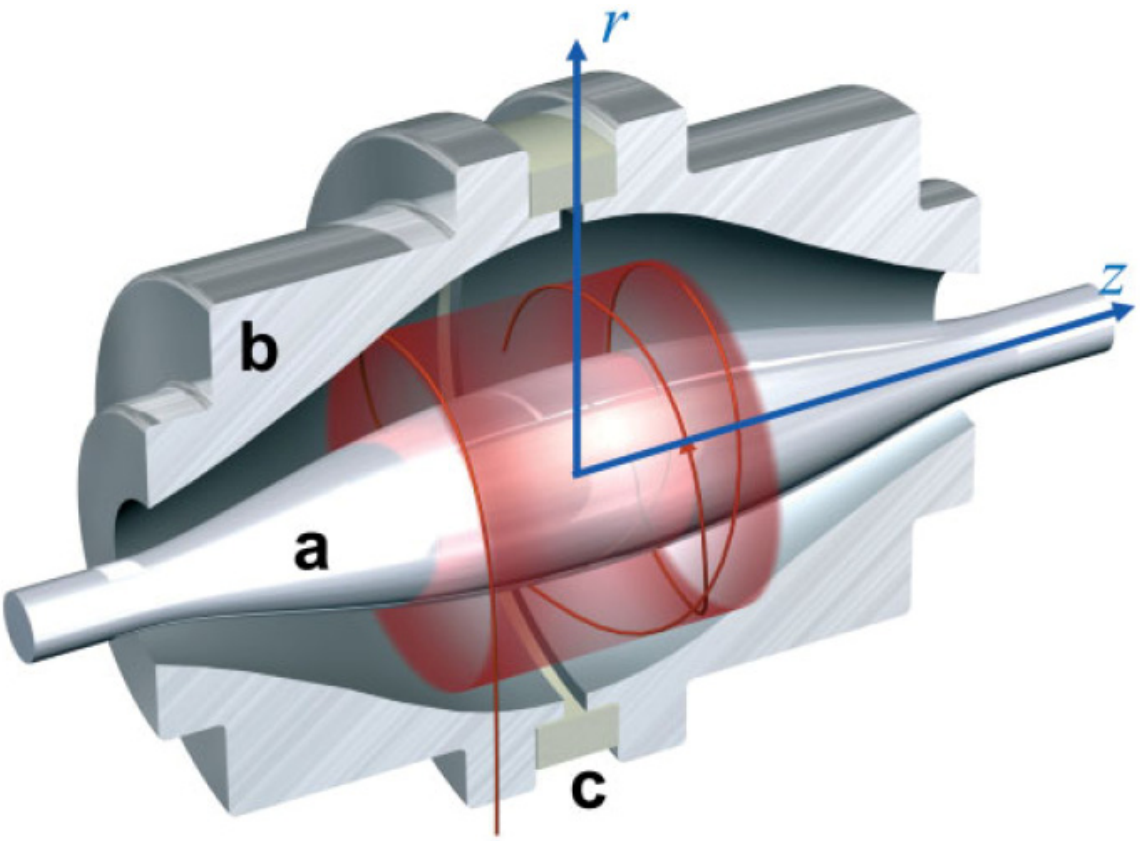


Figure 1.5: A cut-away model of the orbitrap mass analyzer[119]. Ions are moving in spirals around a central electrode (a). An outer electrode (b) is split in half by an insulating ceramic ring (c). An image current induced by moving ions is detected via a differential amplifier between the two halves of the outer orbitrap electrode. The *m/z* ratio of different ions in the orbitrap can be determined from respective frequencies of oscillation after a FT.

1.2.2.4 Quantification

In some cases, the objective of the proteomic study is to compare samples (e.g. normal vs disease) in their respective protein composition. Unfortunately, only slight changes in experimental conditions or sample composition can significantly affect the retention, ionization and detection of peptides and proteins in LC-MS, making comparison between runs very difficult. In those cases, tools have been developed to perform relative quantification between samples. This strategy involves several steps, and usually begins with the labeling of the proteins in the samples with chemically identical but isotopically different reagents followed by pooling and digestion of the samples for simultaneous LC-MS analysis. Identical peptides from both samples will elute together and differ in mass only by the difference in their tag, allowing comparison of the elution peaks. The most frequent represention of this approach is isotope-coded affinity tagging, or ICAT^{TM,121}. The ICAT^{TM} reagent contains a cysteine-reactive group, a linker containing light (8 hydrogens, d0) or heavy (8 deuterium, d8) isotopes and a biotin affinity tag. The proteins are labeled, enzymatically digested, the peptides are selectively retrieved and LC-MS is performed. The isotopically tagged digested peptides give quantitative MS analysis based on the relative peak intensities/areas of d0- and d8-labeled peptides[121]. The limitation of this costly approach is that only cysteine-containing peptides can be compared. Other conceptually similar methods are employed for relative quantification in proteomics. In stable isotope labeling by amino acid substitution in cell culture (SILAC), isotope-tagged amino acids are incorporated in the proteins during cell growth and the differences in masses between tagged and non-tagged, or between differentially tagged samples can be measured. This strategy was employed in an elegant study of the nucleoli proteome in response to different metabolic inhibitors[122], which is

further described in section 1.3.1.1. Isotope tags for relative and absolute quantification (iTRAQ™) is similar to ICAT™, but in this case four isobaric tags are employed, allowing multiplexing of four samples. Even if they have identical molecular weights, the differential position of the isotopes within the tags will cause MS/MS fragmentation to generate four signals of different mass, allowing the experimenter to analyze the contribution of each sample to the eluted peptide peak[123]. A recent study has compared DIGE, ICAT™ and iTRAQ™ techniques and they were found to provide complementary proteomic information[124].

1.2.2.5 Bioinformatics

The vast complexity of proteomic samples and of the ensuing data is calling for an accelerated development of bioinformatics tools in order to extract relevant information from the experimental results. Although it is beyond the scope of this introduction to review all available softwares for data management, it is still essential to devote some time to the tools that translate data into usable biological information, focusing primarily on protein identification and prediction.

1.2.2.5.1 Protein identification

In the case of simple samples, identification of the protein(s) is done by peptide mass fingerprinting (PMF). In the PMF analysis, the experimental spectrum of peptide masses is compared with theoretical ones computed from protein sequences stored in databases and digested *in silico* using the same cleavage specificity as that of the protease employed in the experiment[125]. A score is attributed to a protein match depending on the number and identity of experimental peptides that can be matched to the theoretical

digest, and since this process is not all or none, a list of possible matches is generated according to their score. Consequently, the scoring process is the key to protein identification in this case. Various scoring schemes have been implemented in multiple algorithms which are part of available softwares. Among the most popular ones are those based on the MOWSE[126] scoring algorithm, such as MASCOT[8], MS-Fit[127] and Profound[128].

However, when dealing with complex samples such as those in subcellular proteomics, information provided by the peptides alone is not sufficient to identify and discriminate all the proteins present. This is why MS/MS is important. An additional spectrum for the MS/MS fragments of each peptide can be generated, allowing the discrimination and deduction of the amino acid sequence of peptides of identical (isobaric) or very similar mass. This is accomplished by peptide fragmentation fingerprinting (PFF). The PFF approach is very similar to the PMF approach, with the difference that it is applied to MS/MS spectra, and hence correlates peptide spectra with theoretical peptides from a database[125]. Once again, the most similar peptide candidates are scored and ranked accordingly depending on various factors, including the number and quality of the fragments matched. All the information can be integrated and peptides originating from the same protein can be grouped, leading to confident protein identification despite low coverage or sample complexity. However, MS/MS identification is subject to numerous complications, among which incomplete proteolytic digestion, bad quality of spectra, inaccurate precursor mass and the presence of PTMs. Several softwares are available to manage PFF-based identification while attempting to circumvent these issues. Among those, MASCOT[8], SEQUEST[129] and X!Tandem[130] are the most widely used, although they obviously all have their limitations. Other major

determinants in protein identification are necessarily the databases interrogated, which are numerous for many organisms such as human[131] and vary greatly in terms of completeness, accuracy, redundancy and annotation. Among the most frequently employed databases, we denote the National Center for Biotechnology Information's (NCBI) Entrez Protein database, the NCBI Reference Sequence (RefSeq) database, and UniProt (consisting of Swiss-Prot and its supplement TrEMBL)[132]. While searching for large databases increases the chance of identifying less common peptides or proteins such as polymorphisms, it also generates more false positives and redundancies, mainly due to the fact that these databases are poorly annotated and contain many redundant sequences such as partial mRNAs and sequencing errors[132]. As a consequence, in most cases the experimenter must manually verify each protein and peptide identification, which is labor-intensive and very time-consuming work. For example, many proteins identified as "unknown" are usually a redundant sequence from a very well defined protein, as easily identified by using the BLAST algorithm[133]. Still at this point, a balance must be chosen between speed of analysis and the capacity to carry out complete online and offline protein identification. A complete review of the methods and caveats in protein identification by MS/MS and database searching is available[132].

1.2.2.5.2 Prediction softwares

Another group of interesting bioinformatics tools are the softwares involved in predicting biological information from the identified proteome. For example, not only is this important to identify all the proteins in an organelle, but it is subsequently very relevant to be able to assess which ones interact directly or indirectly with each other. Using a complex combination of statistical and topological descriptors, tools were

developed to predict interactions between individual proteins as well as between protein interaction networks, without using mRNA expression levels and genetic interactions that can therefore be used for validation[134, 135]. Such analysis was very important in gaining functional significance out of the *Drosophila melanogaster* phagosome proteome described in section 1.3.1.6[67]. Analysis of protein interactions has emerged as a field of its own, known as interactomics[136]. Also, bioinformatics is working on predicting other important features of proteins relevant to proteomics. Efforts are devoted to the prediction of protein structures[137], functions[138], PTMs[139] and subcellular localization[139, 140]. The ultimate challenge finally resides in the integration of all this data and information into *in silico*, whole cell models in a systems biology approach[141].

1.2.2.6 Validation

Just as it is the case with any biological technique, validation of the results obtained in proteomics and subcellular proteomics is of paramount importance. Most modern large scale proteomic studies aiming at producing the complete protein inventory of a system will include validation of at least a few of the most interesting identified candidates. In bottom-up or shotgun proteomics, the process of validation is as important as it is challenging, mainly due to the huge numbers of proteins involved. The first and foremost validation step generally occurs during protein identification. As mentioned above, the bioinformatics tools translating MS/MS data into protein lists employ a series of scoring schemes to maximize the probability that the experimental data is accurately represented in the final results. The better the quality of the MS/MS spectra, the greater the number of peptides identified and the higher the sequence coverage of a protein is, the higher the probability that it was indeed present in the sample. Moreover, independent

statistical models exist to analyze the accuracy of peptide identification[142]. Once high confidence identification is reached, classical biochemical tools must take over. And as it is the case for any validation, one or a combination of several independent methods must be employed.

Among the ideal tools for validation is the generation of knockout animals to evaluate if deletion of the gene coding for the protein of interest affects the morphology or functionality of the subcellular compartment analyzed. However, this type of study is time- and resource-consuming, and not applicable for efficient validation of hundreds or thousands of candidates. Another possibility resides in inactivating the protein expression by using RNA interference (RNAi) which now represents one of the most used technologies[143]. Briefly, RNAi is a powerful natural cellular process in which double-stranded RNAs (dsRNAs) target homologous mRNA transcripts for degradation[144]. Effects of the degradation of a mRNA (and the consequent absence of a protein) can then be assessed. Unfortunately, since it is again expensive, time-consuming and labor-intensive, it is usually not realistic to perform RNAi on such a large scale. For that reason, more accessible biochemical validation tools are usually used. Immunocytochemistry will usually reveal the presence and subcellular localization of a protein candidate, especially in conjunction with well accepted and defined subcellular markers[145]. In cases where the protein investigated exhibits a characteristic or unique function, such as enzymatic activity, this can be assessed in cell or organelle extracts or fractions to corroborate other data. Also, association of specific diseases linked to an organelle's function with mutations in newly identified proteins is a strong validity indicator. Evidently, when possible, complete structural and functional characterization of interesting candidates is desirable to extract as much biologically relevant information as possible. In summary,

even tough no universal validation methods is perfect, the combination of bioinformatics and biochemical tools in proteomics can lead to identification of proteins with high confidence.

1.3 SUBCELLULAR PROTEOMICS

Benefiting from significant advances in MS, the subfield of subcellular or organellar proteomics has recently emerged in order to push the understanding of the cell at a higher resolution. Whereas general proteomes, even targeting one cell type, were an important and necessary extension of the various genome projects, it also became clear that obtaining data from which actual biological processes could be derived would require a yet higher level of refinement in proteomics. Effectively, the subcellular localization of a protein is the microenvironment in which it acts and only proteins permanently or transiently part of the same microenvironment can directly interact. Because classical, density based methods for subcellular fractionation were available and had proven effective since the 1950s, this field proposed to combine them to the most up-to-date proteomics methods and analyze the proteomes of the individual subcellular compartments. We will therefore review recent developments in this field, focussing on the proteomes of several organelles and the conclusions that were derived from this information.

1.3.1 Subcellular compartments

Subcellular proteomics deals with small portions of the cell, including organelles themselves and other defined structures such as membranes or the cytoskeleton. In very

broad terms, an organelle is a small membrane-bound compartment presenting a distinct structure and executing precise and defined functions. It encompasses molecules required for its functioning. However, an organelle is not a static entity. Not only does it contain resident molecules that spend most or all of their biological life in the organelle, but it also contains molecules that only transit throught the compartment in order to reach another one or to accomplish a specific function. Since proteins are the most important effectors in a cell, identifying the protein complement of organelles is an efficient way to get important insights on their role in normal as well as pathological conditions. Initially working on complete tissues, proteomics is now zooming in on intracellular organelles in its quest for cell description. Since proteomics-based technologies are getting increasingly accessible, the literature in the field is growing exponentially. We will insist herein on describing recent and thorough studies of selected organelles, some depicted in Figure 1.6, as well as describing some of the important conclusions and insights that were gained from the proteomics analysis of these subcellular structures.

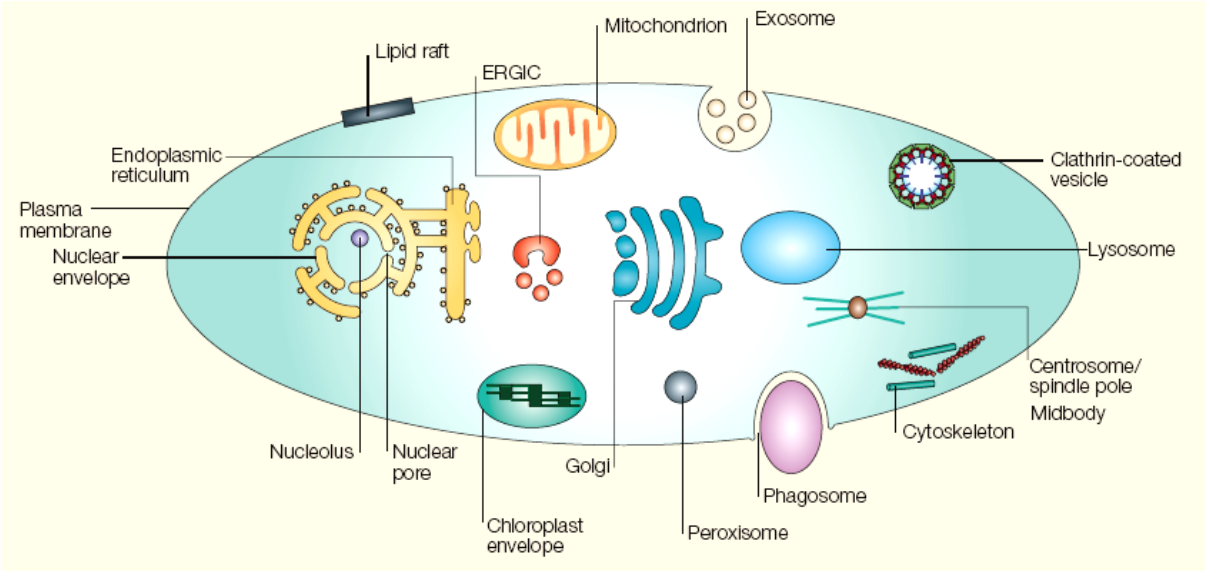


Figure 1.6: Illustration of subcellular compartments that are targets for subcellular proteomics[60]. Not all presented structures are present in every cell, and not all investigated structures are illustrated. For example, dense-core secretory granules (SGs) and synaptosomes found in endocrine and neuronal cells respectively are not shown. Landmark studies for some of these structures are discussed in this section.

1.3.1.1 Nucleus

In eukaryotes, the nucleus is the complex organelle containing the DNA of the cell. It occupies about 10% of the cell volume and is delimited by a nuclear envelope, formed by two concentric lipid bilayers periodically punctured by large nuclear pores responsible for transporting molecules between the nucleus and the cytosol[1]. It is the major location for mRNA synthesis which is then exported out of the nucleus to be translated in the adjacent endoplasmic reticulum (ER). The DNA within the nucleus is highly compacted and associated with proteins in the form of chromatin. The nucleus is of tremendous interest as cancers originate from alterations in the structure or expression of DNA. Proteomics has investigated the two most defined structures of this organelle, the envelope and the nucleolus.

The nuclear envelope (NE) is composed of two adjacent, concentric lipid bilayers separated by approximately 50 nm. The outer membrane is continuous with the ER and also studded with ribosomes. The fact that mutations in a small group of NE proteins is responsible for 15 inherited human diseases with very widespread and apparently poorly related phenotypes has generated interest in understanding the protein composition of this structure[146]. This is especially true since none of the available models at the time could explain the phenotypes[147]. Up until 2001, only about 10 nuclear envelope transmembrane proteins (NETs) were known, and the lack of positive identification of many protein spots on 2D gels suggested that many NETs were still unknown [148]. However, in 2003, a very complete study combining the most up to date proteomic tools was performed and paved the way to rapid progress. In this study, the authors prepared separately NEs and other organelles known to copurify with NE in subcellular fractionation, and *in silico* substracted the proteins in the latter from the NE proteome to increase the probability of

the identified proteins being endogenous to the NE. Using MudPIT, they were able to identify over 80 NE proteins, including 13 known and 67 putative ones with predicted transmembrane domains[149]. Using tagged constructs, they also demonstrated that 8 of the proteins were effectively localized to the NE, thereby validating their approach. Then, using a "guilt by association" approach relating known disease characteristics and the causative effects of mutations in single NE proteins, some of the novel proteins were quickly identified as candidates or causal agents for diseases such as torsinA in early onset torsion dystonia[150].

The nucleolus is the other compartment of the nucleus that has been extensively studied by proteomics. Easily distinguishable under light microscopy, the nucleolus is not membrane-bound, but rather represents a large defined aggregate within the nucleus. Regarded as the structure mainly responsible for producing ribosomes, it encompasses all molecules required for this task including rRNA genes, rRNAs, rRNA processing enzymes and immature or partly assembled ribosomal subunit proteins themselves[1]. Beside its ribosomal production role, recent data suggest that the nucleolus also plays an important role in cell-cycle regulation, senescence and stress responses[122]. An elegant study aiming at analyzing the nucleolar proteome dynamics demonstrates the integration of numerous proteomic methods in deciphering that specific proteome[122]. Not only has this group identified nearly 700 nucleolar proteins, but they have also studied in a quantitative and temporal manner how the endogenous proteins behave in response to three metabolic inhibitors affecting nucleolar morphology. In addition to classical tools, this study employed stable isotope labeling for quantitiation (SILAC) and *in-vivo* fluorescent imaging to visualize the changes that have been measured in the proteomes. Combining this information to that collected from two previous nucleolar proteomic

studies having identified 210 and 257 proteins respectively[151, 152] allows now to draw important conclusions on the biological role of the nucleolus. Firstly, it was discovered that no less that 170 proteins were directly involved in ribosome synthesis, and analysis of the data collected in these three studies resulted in the most complete model to date for that process[153]. Another important result concerns the large proportion of nucleolar proteins involved in mRNA processing, more than would be expected from simple ribosomal synthesis, suggesting that this structure could be an assembly site for translational machinery[153]. Finally, the nature of proteins identified tends to further confirm or corroborate the hypothesized roles of the nucleoli in the regulation of cell cycle[154], apoptosis[155] and aging[156].

1.3.1.2 Endoplasmic reticulum

The endoplasmic reticulum (ER) is mostly described as a large, membranous network of interconnected tubes and flattened sacs that can be further subdivided into rough (with membrane-bound ribosomes) and smooth (no ribosome) regions. This compartment projects from the nucleus into the cytoplasm and, in mammalian cell, can occupy about 10% of the total cell volume and account for 50% of total cellular membranes[1]. It plays crucial roles in protein and lipid biosynthesis. Proteins targeted to various organelles in the secretory pathway usually possess a specific signal sequence and are first co-translationally inserted into the rough ER lumen through specialized translocators. The ER lumen contains numerous enzymes and chaperones whose roles include assisting the nascent proteins to fold properly and ensuring quality control. This organelle will also contain many transient proteins that must first undergo proper processing before continuing their trip toward other subcellular compartments. As it

represents a key mandatory regulatory checkpoint for many proteins, the ER is subject to a great deal of interest in proteomics.

Until very recently, one of the most important attempt in characterizing the ER luminal proteome had already led to the identification of 141 individual proteins[157], confirming that this organelle was amenable to proteomics. Combining subcellular fractionation of murine livers, 2D-gel electrophoresis and MALDI-MS/MS had been successful in identifying 6 new proteins, including ERp19 and ERp46, two new protein disulfide isomerases. However, in late 2006, a very thorough quantitative study of the secretory pathway proteome of hepatic cells provided much more information on the ER[158]. Even if subcellular fractionation in this case still relied on cell lysis and rounds of density-based centrifugation, extreme care was taken to assess the purity of the different fractions. This was accomplished by both biochemical assays for organelle-specific enzymes, quantitative western blotting for accepted organellar marker proteins as well as electron microscopy. The authors estimate that cross-contamination from other organelles in each fraction, still one of the main caveats in this type of studies, was minimized to less than 20%, but was still significant. In total 832 proteins unique to this organelle were identified, and were classified as ribosomal proteins, translocon constituents, molecular chaperones, and proteins involved in lipid oxidation and drug detoxification, as well as proteasome subunits and ubiquitin. Moreover, 405 additional proteins common to the ER and Golgi apparatus were also identified, and were either mainly ER-associated, Golgi-associated or expected to travel in the vesicles between the two, the compartment known as the ER to Golgi intermediate compartment (ERGIC[159]). In order to reach further resolution, smooth and rough membrane fractions were submitted to high salt washes and triton extraction.

Aside from providing a proteome of the early secretory pathway that could be near complete as deduced by annotations and bioinformatics predictions[160], it leads to other important conclusions as far as ER is concerned. First, it extended the known ER proteome by 35%. Secondly, it showed the validity of MS peptide counting as a measure for quantification and contamination. Finally, it indicated that ER proteins extracted in the same fractions (extraction conditions) also clustered in groups of proteins from the same functional class. Six such clusters were defined: protein synthesis and folding, cytoskeleton, metabolism, detoxification and protein modification. In addition to highlighting the important roles of the ER, it suggests that biochemical fractionation in this case led to the specific enrichment of proteins with specific suborganellar locations[158]. Whether this is due to protein-protein or protein-lipid interations is not certain, although this kind of distribution and clustering could help identifying the role of unkown and previously uncharacterized proteins in an associative manner.

1.3.1.3 Golgi apparatus

The next major compartment in the secretory pathway is the Golgi apparatus. It consists of a series of flattened, membrane-bound cisternae loosely organized in stacks. Each stack has two distinct faces, the *cis* (entry) face and the *trans* (exit) face. The Golgi apparatus plays numerous and crucial roles in cell biology. One of these roles is protein post-translational modification, including glycosylation. This modification is accomplished within the Golgi cisternae by the serial and sequential action of various resident glycosyltransferases and glycosidases, resulting in the wide diversity observed in the composition of the polysaccharide groups. Another important function of the Golgi is its role as a protein trafficking regulator. Indeed, it is the final checkpoint from which

newly ER synthesized proteins are sorted to their next (and usually final) subcellular localization, being either, the lysosome, the secretory granule or the plasma membrane to name a few. Moreover, recent evidences suggest a role of this organelle as a scaffold for intracellular signaling[161]. Early 2D-electrophoresis work suggested that over 1000 proteins could be potentially required for this organelle to maintain its structure and accomplish its roles[162].

Interestingly, focussing on the progress in this area clearly illustrates the improvements and proteomics approaches and technologies. In 2000, a study comparing functional states of the Golgi and combining gradient-based subcellular fractionation, 2D-gel electrophoresis and MS/MS led to the identification of 35 proteins[163]. In 2004, the same group reported the use of MudPIT coupled to MS/MS to identify 421 proteins in that organelle. This latter study not only identified 41 new proteins of unknown function, but it also identified arginine dimethylation as a new post-translation modification in Golgi proteins[164]. Additional progress was soon reported and resulted in the further identification of no less than 700 proteins[158]. Moreover, a proteome of over 480 proteins was obtained for COPI-coated vesicles. These structures are though to be responsible for transport within the Golgi and between the stacks, although their role is still controversial[165]. In this case, the results concurred with existing models concerning the formation and role of COPI coated-vesicles, confirming that their proteome could be viewed as an extension of that of the Golgi. Finally, a very recent study led to the identification of 1125 proteins belonging to the Golgi stacks[166]. Once more, the sample was obtained by density gradient centrifugation and the digested peptides were analyzed by multidimensional nanoLC-ESI-MS/MS. Two different softwares were employed for protein identification, and *in silico* analysis was performed to assign an intracellular

localization to some of the unknown proteins. The study of two of these proteins was pushed further, and transfection of HeLa cells with enhanced green fluorescent protein (EGFP)-tagged constructs confirmed their Golgi localization. As a whole these studies not only provide an increasingly accurate and complete representation of the protein content of the Golgi apparatus, but also demonstrate that completely new proteins can be identified and characterized, and that proteomics can contribute to the elaboration or confirmation of relevant models for biological processes.

1.3.1.4 Mitochondrion

Although they are quite dynamic in structure, mitochondria are usually cylindrical in shape, ranging from 0.5 to 1 μm in size. From the outside to the inside, they are composed of an outer membrane, an narrow intervening space, an inner membrane and an internal matrix. Mitochondria differ from other organelles by containing their own circular genome coding for a few of their constituent proteins, while other proteins come from normal nuclear origin. This heterogeneous composition is believed to be responsible for the range of biochemical roles that can be fulfilled by this organelle, ranging from a significant biosynthetic role (liver) to a primarily energy metabolism-oriented organelle (heart)[167]. However, the mitochondria are still viewed as a the major energy producer in the cell, generating ATP from ADP and oxygen by oxidative phosphorylation through the respiratory chain, a group of oxidizing enzymes located in the intermembrane space.

Precursor large-scale proteomic studies had provided helpful information in attributing 546, 591, and 615 proteins to yeast[168], mouse[169] and human[170] mitochondria respectively. However, based on a human genome analysis, there could be as much as 3000 proteins localized to this organelle[171]. A recent study has markedly improved the

mitochondrial proteome coverage. Comparing the protein composition of this organelle in four different rat tissues, brain, liver, heart and kidney has allowed the identification of 1162 proteins, including 145 that, by virtue of their N-terminal mitochondrial localization targeting sequences[172], were associated with this organelle for the first time[167]. Moreover, 1149 of those protein were shown to be significantly different in abundance between tissues using a mass spectrometry quantification software[173]. This correlates well with the variability in physiological function of the mitochondria and led the authors to further confirm the notion that mitochondria are tuned by the nucleus for specific functions in different tissues[167]. Also, numerous comparative, rather than descriptive, proteomic studies of the mitochondria have been accomplished in order to identify protein markers responsible for functional changes in this organelle. The first example of such studies compared the mitochondria proteomes in normal gastric cell line and tumor gastric cell line in the quest for biomarkers for the transformed phenotype[174]. Protein were separated by 2D-gel electrophoresis and identified by MALDI-MS, and 14 spots presented very significant differences between the two samples. Interestingly, most of these proteins could be categorized in cancer-related proteins, stress-related-proteins and chaperonins, and energy metabolism-related proteins. The authors relate these changes to the adaptation of the cell to hypoxia and increased metabolic state. Another group investigated the effect of aging on rat brain tissues through proteomics using similar techniques[175]. Here again, age-specific differences were observed in individual proteins as well as large complexes involved in oxidative phosphorylation, hinting a role for these in the alterations in reactive oxygen species production during aging. Since the free radical theory of aging postulates that the production of intracellular reactive oxygen species is a

major determinant of life span[176], identification of proteins involved in their production and management could provide crucial new information on this process.

1.3.1.5 Lysosome

The lysosomes can be quite variable in size and shape and are the main intracellular degradative compartment. The acidic lumen of the lysosomes contain a battery of acid hydrolases whose main roles are to perform a wide variety of digestive functions, including the breakdown of intra- and extracellular debris, the destruction of phagocytosed organisms and the production of nutrients for the cell[1].

The luminal enzymes of this organelle have been extensively described by classical and proteomics studies[177] and defects in some of these enzymes can lead to serious diseases[178]. However, much work still need to be performed on the membrane proteins of this organelle. Two recent studies have addressed this issue. In the first one, proteomics analysis of triton-filled lysosomes, termed tritosomes, led to the identification of 215 proteins from a membrane enriched fraction[179]. Among the interesting findings it is worth noting the presence of significant amounts of ER and cargo proteins, as well as the abundance of lipid raft proteins, confirming the presence of this structure in this organelle. More recently, the secretory lysosomes found in natural killer (NK) cells were the focus of attention. In such specialized immune cells, this organelle performs two distinct functions: namely acting as a degradative compartment and also acting as a secretory organelle that delivers perforin and granzymes to the immunological synapse[180, 181]. Combining subcellular fractionation, triton-based extraction and enrichment of membrane proteins, SDS-PAGE and LC-ESI-MS/MS, the authors have identified 221 proteins, of which an impressive 61% were predicted to be integral membrane or

membrane-associated proteins[181]. Many of the identified proteins belonged to the Rab GTPase and soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptors (SNAREs) families known to be intimately associated with membrane fusion processes. More specifically, their evidence reinforced important roles for Rab27a[182] and VAMP7[183] as potential key components of the machinery that regulates the exocytosis of the secretory lysosomes in NK cells, shedding some light on the mechanisms by which cytotoxic T lymphocytes and natural killer cells eliminate their targets.

1.3.1.6 Phagosome

Phagocytosis is the specialized endocytic process by which cells can ingest large particules such as microorganisms. Derived from protozoa and early eukaryotes in which it was essentially employed as a method of feeding, phagocytosis has evolved into a complex and essential immune process to protect the organism from pathogens. The size and morphology of the phagosome depend largely on the endocytosed particule and are therefore quite variable. The phagosome is viewed as comprising three distinct compartments: the pathogen itself, the luminal content including hydrolases, protons and ions, and the membrane, the only boundary between the cell and the pathogen[184].

A unique character of the proteomic studies of phagosomes distinct from that of other subcellular structure relies in the fact that they utilize a significantly modified subcellular fractionation procedure involving a single step but which leads to high levels of purity. It was known for over three decades that phagosomes having enclosed latex beads could be rapidly and cleanly isolated due to the fact that they float in a sucrose gradient[66]. A pioneering study employing this strategy for phagosome purification coupled to 2D-gel electrophoresis and MS led in 2001 to the identification of over 140

proteins, among which unexpected candidates exhibiting pro- and anti-apoptotic roles[185]. More than simply allowing elaboration of a model for the phagosome, the presence of ER proteins in the list led to the hypothesis that some of the phagosome and lysosome membranes could originate from the ER[186]. Another recent and impressive study of the *Drosophila melanogaster* phagosome using a systems biology approach has provided the most complete understanding of this organelle to date[67]. Combining the above-mentioned latex incorporation-based purification, gel electrophoresis and MS/MS, 617 phagosome proteins were identified, including 122 transmembrane and 107 previously undefined proteins. Using bioinformatics tools[134, 187], the authors established the networks of interactions of these proteins between themselves and with others (interactome) in order to gain more substantial biological information out of their data. Last but not least, using RNAi technology to selectively knock down the 617 identified protein and an additional 220 which could interact with them, they assessed their role in the phagocytosis of Gram positive and negative bacteria. The fact that 28% of the proteins influenced either positively or negatively the phagocytic process is much in favour of the systems biology approach and highlights the power of this kind of analysis in that it gives rise to a rich spectrum of molecular and testable hypotheses that can drive the entire field[184].

1.3.1.7 Synaptic vesicle

Another organelle that has captured the researchers' interest for several decades is the synaptic vesicle. Being quite small (50 nm) and homogeneous in size, synaptic vesicles are very abundant in the presynaptic terminals of neurons. The role of these vesicles is to store small molecule neurotransmitters and mediate their regulated release at synapses to effect rapid cell to cell signalling. Aside from their neuronal function, these

small synaptic vesicles are also considered an excellent and relatively simple biological model for membrane trafficking[188]. Consequently and until very recently, most proteomics studies focussed on the synaptic vesicle membrane proteins.

In 2004, two complementary 2D gel eletrophoresis methods combined with MS led to the identification of 36 proteins, including 7 integral membrane proteins[189]. A year later, 72 proteins were identified by 2D-gel electrophoresis (BAC (Benzyldimethyl-n-hexadecylammonium chloride)- in the 1st dimension, SDS- in the 2nd dimension) and MS[65]. In 2006, analysis of the synaptic vesicle proteome using three gel-based protein separation techniques and MS to maximize the number of membrane proteins identified resulted in a proteome of 185 proteins. The identified synaptic vesicle proteins included transporters, SNAREs, synapsins, Rabs and Rab-interacting proteins, additional guanine nucleotide triphosphate (GTP) binding proteins, cytoskeletal proteins, and proteins modulating synaptic vesicle exo- and endocytosis from which a model could be derived[190]. Moreover, it illustrated the difficulty of performing gel electrophoresis on membrane proteins since very few of the proteins could be identified by all three methods. Finally, a thorough study published recently has succeeded in characterizing to an unprecedented level the complete anatomy of the synaptic vesicle in a systems biology perspective not too distant from that which inspired the *Drosophila* phagosome analysis described earlier[188]. This group has employed again BAC/SDS-PAGE and 1D-PAGE followed by MS to identify 410 synaptic vesicle proteins, including over 80 integral membrane proteins. The abundance and diversity of trafficking proteins were surprising, since over 20 different SNAREs proteins were detected. The presence of selected proteins was also quantified using Western and dot blotting, and the location of others was confirmed by immunogold staining. Other physical properties of the vesicle were

assessed, including size and lipid composition of the membrane. With this information in hands, the authors were able to derive a model for the vesicle, shown in Figure 1.7, and the possible implications of this model to membrane structure, membrane trafficking and neurotransmitter release and uptake[188]. These implications were further discussed[191] while recent progress in the field was reviewed[192].

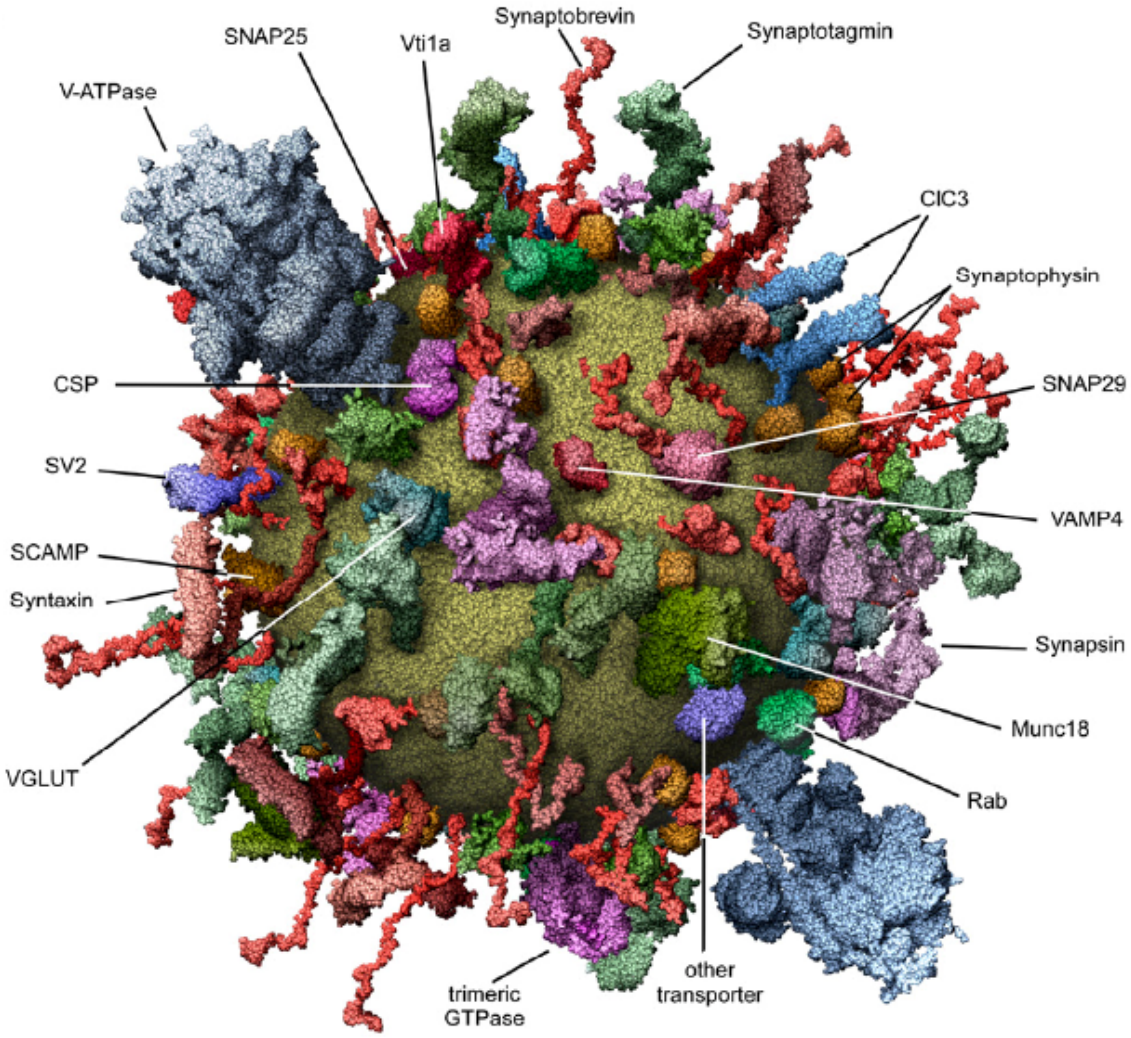


Figure 1.7: Molecular model of the synaptic vesicle[188]. The model is based on space-filling models of all macromolecules at near atomic resolution, and includes proteins identified in the cited proteomic studies as well as others previously reported.

1.3.1.8 Dense-core secretory granule

The dense-core secretory granule (or secretory vesicle) (SG) is a specialized organelle present in endocrine and neuroendocrine cells. Originating from the *trans*-face of the Golgi, SGs undergo progressive maturation as they travel toward the plasma membrane where they will remain until the cell receives the appropriate stimuli for them to release their cargo outside the cell. Dense-core SGs are characterized by the presence of a dense protein aggregate core surrounded by a membrane. Their main role is to store, transport and secrete proteins destined to be released outside the cells in a regulated manner. The structure and specific functions of this organelle will be thoroughly discussed in section 1.4.3.

The SG is an organelle that has been surprisingly neglected as far as proteomic effort is concerned. Until early this year, not a single large scale proteomic effort had been published on this structure. For example, only approximately 30 insulin-containing SG proteins were known, even though early 2D-electrophoresis data suggested that over 150 polypeptides could be present[193]. In the last few months, studies on two types of endocrine secretory granules were reported. In the first one, insulin containing SGs were first purified from INS-1 cells using standard subcellular fractionation procedure including density-based centrifugation. The protein extract from purified SGs was separated by 1D-gel electrophoresis and the proteins bands were identified by LC-ESI-MS/MS. In this context, 130 proteins were identified, including 110 that were new to the secretory SGs[194]. Aside from being the first large-scale proteomics analysis of this organelle, the results presented therein revealed a large number of lysosomal proteins such as hydrolases (cathepsins[195]) as well as membrane proteins (lamps[195]) to be present in the SG preparation. A very similar approach was employed to study another type of

secretory granules, known as chromaffin granules, present in bovine adrenal medulla. These granules are responsible for secreting neurohumoral factors in response to stress and have been long recognized as an excellent model of a secretory organelle[196]. As an additional refinement, the group separated the protein extract into soluble and membrane proteins and performed identification by MS following 1D-gel electrophoresis. In total, 102 chromaffin granule proteins were identified, including 43 that were common to the soluble and membrane pools[197]. Interestingly, very few lysosomal proteins were identified, contrasting sharply with the previously mentioned study. The chromaffin granule proteome was subdivided into 14 distinct functional categories. Aside from the expected prohormones and prohormone processing enzymes, proteins were identified to play roles in a much wider range of processes including proton regulation, protein folding, carbohydrate and lipid metabolism and signalling. This information was integrated in the form of a systemic diagram shown in Figure 1.8.

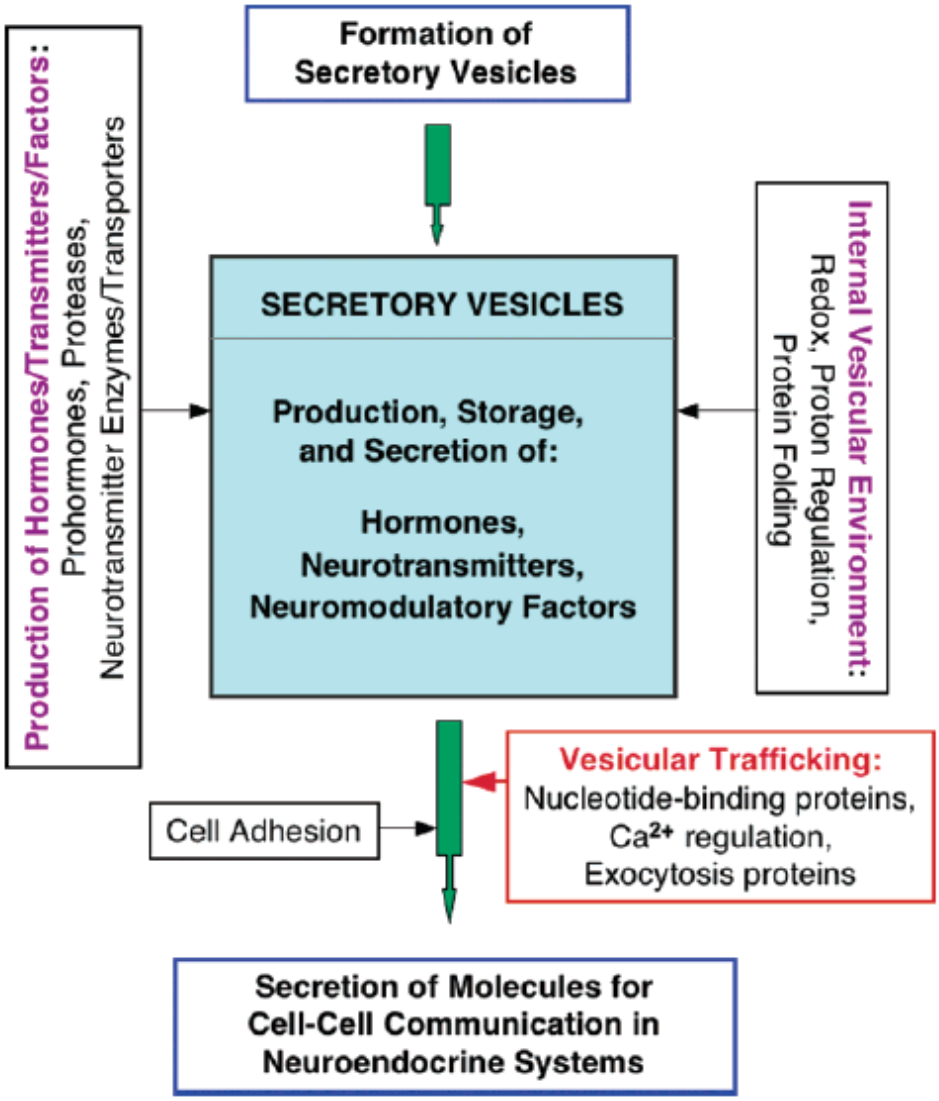


Figure 1.8: **Functional protein systems for secretory vesicle production, storage, and release of peptide hormones, neurotransmitters, and neurohumoral agents**[197]. Among the protein systems identified, we denote those implicated in SG formation, hormone processing and production, establishement of intragranular microenviroment, vesicular trafficking and secretion.

Finally, these two studies illustrated the higher than expected complexity of this specialized organelle's proteome, hinting that more work in this field is essential to close the gap with what has already been done for other organelles. Also, it was a clear indication that, as strongly suggested before, there is a need for alternative and reliable methods for the purification of subcellular structures if accurate interpretation of the proteomics data is to be achieved. The work presented in this thesis will address both of these aspects.

1.3.1.9 Plasma membrane

The cellular plasma membrane is a lipid bilayer encompassing the cellular cytoplasm and all other subcellular structures. This fluid membrane is mainly composed of four types of phospholipids (sphyngomyelin, phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine) and of cholesterol. It acts to separate the intracellular from the extracellular milieus in order to preserve the integrity of the cell and its homeostatic equilibrium. However, in order to exchange both material and information with its environment, the cell uses multiple membrane-associated and transmembrane proteins. These can be divided into numerous functional categories, such as receptors, channels, pores and can be either free floating in the membrane or more strictly organized into membrane loci known as lipid rafts. Upon reception of the appropriate stimuli, these proteins will generally undergo conformational changes allowing them to transmit the biological molecule or the information from one side of the plasma membrane to the other. Because of their accessibility and their important biological roles, membrane proteins are the targets of choice for drug design. Unfortunately, because of their physicochemical properties, they are also usually difficult to study. Indeed, due to their

hydrophobic membrane-spanning domains, these large proteins are rarely soluble in aqueous buffers and are poorly resolved by SDS-PAGE.

The field of proteomics has focussed on developing new tools for membrane proteins enrichment, solubilisation and separation in order to be able to study this important subset of proteins. Such work is well illustrated by the mapping of the lung endothelial surface in normal and malignant tissues. Using silica-coated affinity beads to isolate plasma membranes, the authors have identified over 2000 membrane proteins and validated two of them as being effectively exposed to the surface[198]. Moreover, a substrative proteomics approach allowed the authors to identify 12 proteins that were enriched in tumor endothelium. One of these proteins was annexin A1, and radioimmunotherapy against annexin A1 induced tumor destruction, confirming the validity of this large scale method to identify therapeutic targets for cancer treatment. Another study has taken advantage of the changes in the density of membranous compartments induced by digitonin treatment which can be proportionally related to their content in cholesterol. Using this approach, it was possible to isolate plasma membranes and identify 1685 proteins from mouse hyppocampus[69]. In the last few months, by refining this work and combining this membrane purification protocol with HysTag, their new tool for peptide quantification[199], they have been able to quantitate 976 peptides from 555 membrane proteins in three distinct brain regions. The identified proteins include many interesting glutamate receptors, calcium channel subunits, and ATP-ases[200]. Other recent and practical applications of membrane proteomics include monitoring of cellular differentiation. For example, 137 of 1125 proteins were shown to be specifically and significantly increased in either proliferating or differentiatied human colorectal carcinoma (Caco-2) cells[201]. Most of the proteins increased in the differentiated cells were

metabolic enzymes, proteins involved in the maintenance of cellular structure, transmembrane transporters, and proteins regulating vesicular transport whereas the majority of the proteins increased in the membranes of proliferating cells were involved in gene expression, protein synthesis, and folding[201]. This study provided several new candidates for specific markers of intestinal cells and colorectal tumor. Another remarkable application of membrane proteomics is the search for membrane markers of viral infection. Comparing the proteomes of latently HIV-1 infected cells with the non-infected parental cell lines, the presence of over 10 proteins at the cell surface could be attributed to the infection[202]. Inhibition of two of these proteins, Bruton's tyrosine kinase and X-linked inhibitor of apoptosis, led to an increase in drug treatment susceptibility, suggesting this approach for the discovery of new therapeutic targets.

1.3.1.10 Other organelles

The peroxysome, responsible for performing oxidative reactions including those involved in plasmalogens (abundant phospholipids in myelin) synthesis, was studied. A comparative study between mouse liver and kidney peroxysomes has revealed 91 proteins and led to the elaboration of the first database[203]. Other peroxysome proteomic work has been reviewed[204]. The clathrin-coated vesicles (CCVs), playing roles in transport within the endosomal pathway, was also thoroughly investigated and 209 proteins were identified[205]. This work highlighted the importance of this organelle in synaptic vesicle recycling and identified a novel protein, enthroprotein, linked to schizophrenia[60, 205]. Another recent study on this organelle, though it led to the identification of only 63 proteins, interestingly was based upon an approach aiming to ensure that they were *bona fide* clathrin-coated vesicle proteins as they compared their fraction of interest with a

"mock" CCV fraction originating from clathrin depleted cells[206]. The chloroplast, an organelle related to the mitchondria and responsible for photosynthesis in plants has also been investigated. Significant numbers of both soluble[207] and membrane[208, 209] proteins were identified. Finally, large and less distinct subcellular structures are also amenable to proteomic analysis. For example, the cell cytoskeleton, a complex network of filaments and tubules implicated in determining cell shape and movement was also put under scrutiny. Several studies have isolated and analyzed this proteome[210, 211]. Results on the proteomics of these and other subcellular compartments have been summarized elsewhere[60, 212].

1.3.2 Challenges in subcellular proteomics

Despite its fast-paced progress, subcellular proteomics still faces numerous recent challenges. For example, the amounts and complexity of the computerized data generated calls for improvements in bioinformatics. Improved algorithms for MS-based protein identification from much higher quality databases are required. Also, tools to integrate and analyze the identified proteins into distinct structural and functional categories, as well as in biologically relevant interaction networks will be needed for asking large-scale biological questions. On another hand, some challenges are several decades old, but of equal if not superior importance. And the most important of all is certainly related to the quality and purity of the subcellular structure preparation to be studied, the foundations upon which the "omics" tower is erected. Since validation of the data is becoming an overwhelming and tremendously expensive task, "separation methods that give a higher resolution would facilitate the analysis of all cellular components using proteomic technologies, and so the development of new methods is a challenge for the field"[60].

1.4 SPECIFIC AIMS OF THIS THESIS AND THE BIOLOGICAL MODELS USED

The work presented in this dissertation involves the design, development and application of new methods in subcellular proteomics. In order to optimize and test the new approaches, several biological models were employed. As far as the development of a nondiscriminatory investigative approach for the study of cell surface proteins is concerned, our first feasibility study was performed using Sf9 insect cells. (chapter 2)

The results were sufficiently convincing to apply it then to the study of a subcellular proteome of greater interest to us, the cell surface of the mammalian spermatozoa. (chapter 3)

Our other method consisted in the development of one of the first functional, fast, efficient and adaptable alternative to density gradient centrifugation for organelle purification, namely using FAOS. Using our approach, we have obtained the proteome of the secretory granules of AtT-20 corticotrope cells in less than one day. (chapter 4)

The following text is an introduction to the various biological models that we have used in those studies.

1.4.1 Sf9 insect cell

Sf9 cells are a clone of the IPLB-SF 21 cell line established from the armyworm *Spodoptera frugiperda*[213]. This lepidoptera originates from southern and central America and mainly feeds on corn, sorghum and grasses, presenting a serious threat to the crops [214]. Sf9 is one of the various insect cells used in research as they are relatively cheap to maintain and are capable of producing large amounts of accurately translated and

correctly processed heterologous proteins[215]. The baculovirus expression system is one of the most widely used to express recombinant mammalian proteins in insect larvae and insect cells such as Sf9[216], even though there are some fundamental differences in glycosylation pathways between insects and higher eukaryotes[217]. Sf9 cells are commercially available and have been adapted for growth in serum-free media to simplify recombinant protein purification. There are hundreds of examples of successful mammalian protein production using this model, including work done in our laboratory on the prohormone convertase 1/3 (PC1/3)[218, 219]. Its ease of culturing and its immediate availability made it a very attractive model to develop our approach for identifying membrane proteins. On the other hand, this organism and cell line have been quite poorly studied and so far only 232 protein entries exist in the Entrez global query cross-database search system, including many ribosomal proteins for which redundancies are already present. Hence, obtaining additional information on the proteins present at the surface of Sf9 insect cells could certainly benefit the field of fundamental research and probably identify targets for improving baculovirus-mediated infections. More practical applications of this work could concern the development of more targeted insecticides to counteract the increasing resistance of this insect[220, 221].

The objectives in this study were to i) synthesize and characterize a new biotinylation reagent possessing the physicochemical properties necessary for the labeling of membrane proteins and the easy purification of their labeled peptides, ii) perform a proof-of concept study for our non-discriminatory investigative approach for cell surface-exposed proteins and iii) identify, if possible, new Sf9 surface-exposed proteins.

1.4.2 Spermatozoon

1.4.2.1 Introduction: PC4 and spermatozoon structure

Aside from PC1/3, our laboratory is also interested in another member of the prohormone convertases family, namely PC4. PCs are calcium-dependent serine proteases responsible for the proteolytic maturation of numerous proteins including growth factors, hormones and receptors in physiological and pathological conditions[222]. PC4 was discovered in 1992 by virtue of its homology to other members of the family, cloned and the mRNAs were found to be mainly located in the germ cells of the testes in rat[223]. Several distinct alternate mRNA forms for the gene were identified, and at least one of them codes for a protein presenting a possible GPI-anchoring site at the C-terminus[224]. A mouse model in which the gene was inactivated results in the male being infertile[225]. More recently, experiments using immunocytochemistry, immunoelectron microscopy and immunofluorescence and showed that the enzyme was detected in the acrosomal granules of round spermatids, in the acrosomal ridges of elongated spermatids, and on the spermatozoa plasma membrane overlying the acrosome[226]. It also revealed that sperm from mice genetically deficient for the PC4 protease exhibit accelerated capacitation, precocious acrosome reaction, reduced binding to egg zona pellucida and impaired fertilizing ability, suggesting a role of this surface enzyme in the maturation and function of this cell[226]. For this reason and also because of the fact that sperm surface proteins present a great interest as far as fertilization and contraception are concerned, we have decided to use our newly developed method to define the spermatozoa cell surface proteome.

The structure of the spermatozoon is peculiar and complex. This cell comprises three distinct functional parts, as shown in Figure 1.9, namely the (i) the sperm head,

involved in sperm-oocyte interactions; (ii) the midpiece containing mitochondria, involved in energy production; (iii) the flagellum, involved in motility[227]. The head is further subdivided into four regions, the membrane composition of each being different.

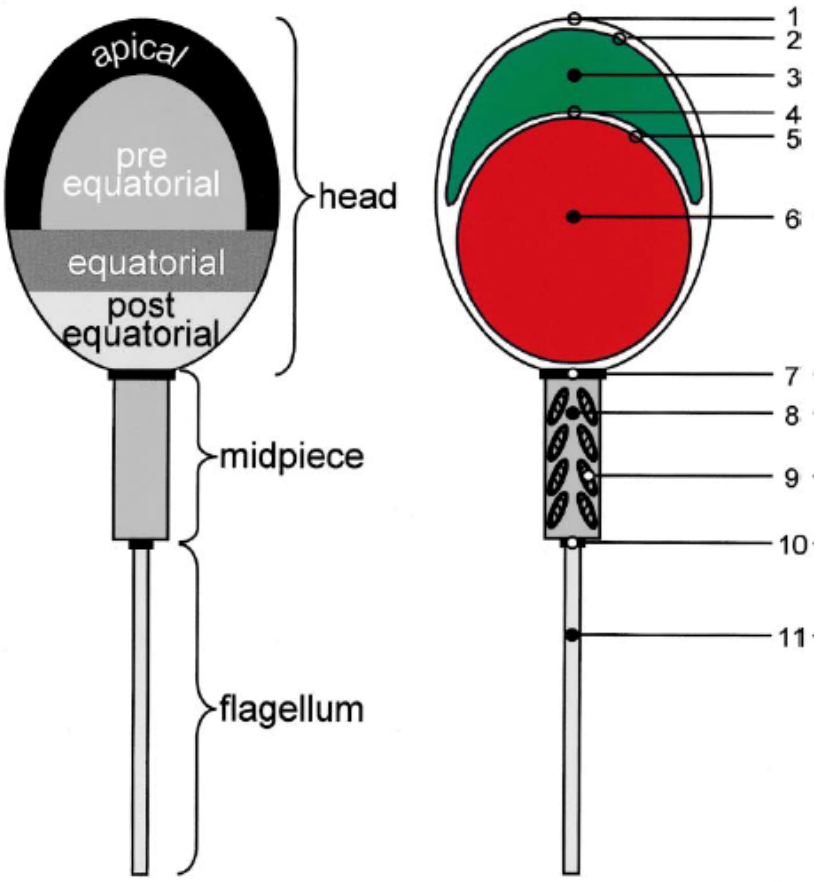


Figure 1.9: Illustration of the important regions and structures of the mammalian spermatozoon[227]. 1: plasma membrane; 2: outer acrosomal membrane; 3: acrosomal content; 4: inner acrosomal membrane; 5: nuclear envelope; 6: nucleus containing highly condensed DNA; 7: posterior ring; 8: midpiece; 9: mitochondrion; 10: annular ring; 11: flagellum.

As a whole, the process of fertilization of the oocyte by the spermatozoa is quite complex and involves a series of sequential changes in the latter. Epididymal or freshly ejaculated spermatozoa are not able to perform fertilization of the oocyte until they spend some time in the female reproductive tract. The molecular and physiological events that confer on the sperm the ability to fertilize during residence in the female tract are collectively known as capacitation[228]. Among other transformations, there are marked changes in the abundance and distribution of membrane proteins among the different head regions. Once the capacitated spermatozoa reach and make contact with the oocyte, they will empty the content of their acrosome, a large vesicle containing hydrolytic enzymes necessary for the penetration of the oocyte extracellular matrix (zona pellucida, ZP), in a process known as acrosome reaction[229]. At this point, the inner acrosomal membrane becomes exposed and the spermatozoa can bind the ZP, penetrate and fuse with the egg. From that whole process, it is apparent that sperm plasma membrane, along with the outer and inner acrosomal membranes, contains interesting surface-exposed proteins critical to fertilization. Acquisition of the cell surface proteome of this cell model could therefore provide extremely important biological information.

1.4.2.2 Cell surface proteins in capacitation

As mentioned above, the process of capacitation induces changes in the sperm membrane that increase its affinity for the ZP[230]. It has been described that tyrosine phosphorylation occurs on several sperm proteins during *in vitro* capacitation, even if this has not yet been clearly confirmed *in vivo*[231]. The three major components responsible for capacitation *in vitro* are bicarbonate[232], albumin[227] and Ca^{2+},[233], although the former is though to be mainly involved in triggering protein tyrosine phosphorylation in

capacitating mammalian sperm. Tyrosine phosphorylation of sperm proteins is linked with increased ZP affnity[234], acrosome reaction[234] and hypermotility[235]. Moreover, capacitation was shown to induce tyrosine phosphorylation of a specific subset of membrane proteins, some showing high affinity for ZP material[236]. It is therefore clear that at least post-translational modification of sperm surface proteins is crucial for capacitation. On another hand, extracellular Ca^{2+} is also required for sperm capacitation and the ability to undergo induced acrosome reaction[237]. Calcium influx through the sperm plasma membrane can occur through the action of either and at least Ca^{2+} channels[238], Ca^{2+} ATPases[239] and Na^{+}/Ca^{2+} exchangers[240] present at the surface. This action is not only regulated by the presence and abundance of such proteins, but also by the regulation of their activity. For example, inhibition of a Na^{+}/Ca^{2+} exchanger by caltrin, a 10 kDa seminal plasma peptide, prevents Ca^{2+} uptake by freshly ejaculated bovine sperm cells[241]. Finally, changes in the glycocalyx, the organized extracellular coating of the sperm surface were observed by lectin binding studies during capacitation. Aside from the demonstrated repositioning of glycolipids[242], several "decapacitation" factors and other proteins are removed from this extracellular coat in this process. This event is proposed to lead to protein tyrosine phosphorylation and to expose important proteins such as the progesterone receptor[243] at the sperm's surface. A recent study using mouse spermatozoa surface biotinylation and *in vitro* capacitation led to the identification and validation of at least four proteins released from the cell's surface, including plasma membrane fatty acid binding protein, cysteine-rich secretory protein 1 (CRISP1), phosphatidylethanolamine binding protein 1 (PBP), and a largely undefined protein product known as decapacitation factor 10 (DF10)[244]. Spermadhesins, major secreted products in the seminal plasma, are other examples of proteins that are found at the

surface of spermatozoa and subsequently removed during capacitation[245]. BSP proteins, a family of phospholipids binding proteins found in the seminal plasma, are also known to bind both the sperm membrane itself and several capacitation factors, thereby promoting capacitation[246].

1.4.2.3 Cell surface proteins in zona pellucida binding and in acrosome reaction

Following capacitation, the spermatozoa are then able to reach and bind the ZP surrounding the oocyte. Sperm binding to the ZP ultimately evokes the acrosome reaction, which is required for successful penetration of the ZP (and thus for fertilization)[227]. This binding occurs in two stages. First, acrosome-intact cells will bind the ZP, which will induce fusion of the plasma membrane with the outer acrosomal membrane, leading to the release of the content during the acrosome reaction. Secondly, the acrosome-reacted spermatozoa will now expose several inner acrosomal membrane proteins which will be involved in this secondary binding to ZP. Even though primary binding could be as much as 75-80% carbohydrate-dependent[247], the protein interactors involved in the remaining 20-25% of the binding are still very poorly characterized. Several proteins candidates exposed at the surface of sperm have been identified as being potentially involved in the primary binding[248]. Among the most interesting proteins are a β-1,4 galactosyl transferase[249], a 56 kDa protein[250] and a tyrosine phosphorylated protein, p95[251]. The identities of the players involved in primary sperm-ZP binding are still, however, largely undefined. Immediately following primary binding of the ZP, the acrosome reaction is initiated[252]. Once again, evidences suggest that receptor aggregation[253], ion channels[254], progesterone[255] and protein phosphorylation[256] are implicated. Interestingly, G-protein signaling also seems to be an essential reaction as its

inhibition prevents acrosome reaction without affecting binding[257], while its activation induces acrosome reaction[258]. Following acrosome reaction, new intraacrosomal proteins also become exposed. Some of these proteins are involved in secondary sperm-ZP binding and have been more thoroughly studied. Among the candidates identified, we denote P-selectin[259], sp38[260], acrosin[261] and PH-20[262], the latter two possessing a protease and a hyaluronidase enzymatic activity respectively, suggesting a role in both binding to and digestion of the ZP.

Finally, binding to the oocyte plasma membrane (oolemma) is another step in which membrane proteins are very important. Once it has penetrated the ZP, the sperm initially binds the oolemma with the tip of its head, and then through its equatorial segment. Among the proteins involved in those bindings are cyritestin[263] and fertilin (PH-30)[264], two membrane proteins containing a desintegrin and metalloprotease domain, and are therefore members of the ADAM family. Generation of antibodies inhibiting sperm-egg fusion pointed at two other candidates. A novel member of the immunoglobulin type 1 membrane protein superfamily, Izumo, was also found to be crucial for sperm-egg fusion, although its binding partner on the oocyte has not yet been identified[265]. DE (CRISP-1) is a glycoprotein belonging to the cysterine rich secretory proteins family and associates with the spermatozoa membrane during epididymal transit[266]. Figure 1.10 illustrates the current knowledge on the players involved in sperm-egg fusion.

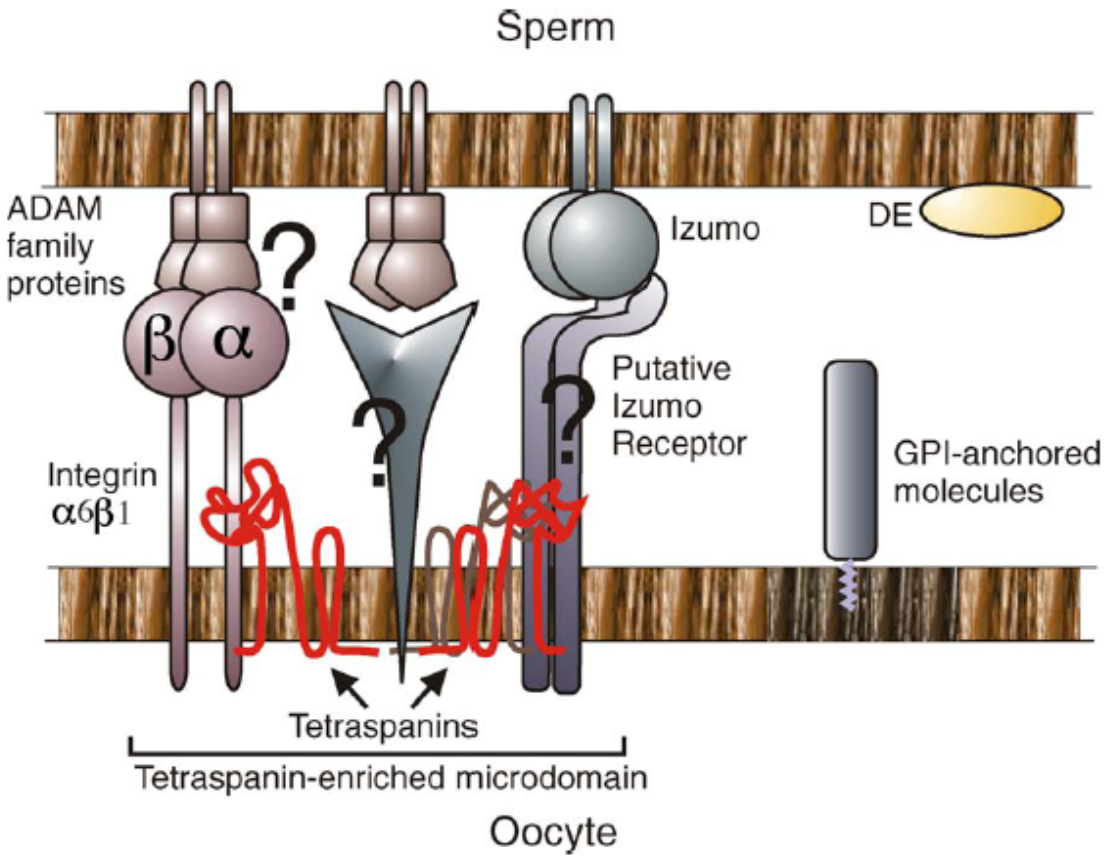


Figure 1.10: Molecules involved in sperm–egg fusion[267]. Integrin α6β1 is a proposed binding partner for fertilin, a member of the ADAM family. The presence of an Izumo receptor in Tetraspanin-rich regions of the oocyte is speculative. Very few interactions between sperm and egg have been undisputedly demonstrated.

Clearly, surface proteins are tremendously interesting and important in the spermatozoon's biology and function and deserve further investigation. Surprisingly however, our current knowledge of those proteins involved in fertilization still contains few certainties. Proteomics, and our specifically designed investigative method for cell surface proteins, would potentially be quite potent in addressing this issue. Spermatozoon is also perceived as a relatively simple biological sample containing only a few hundred major proteins but performing a much wider role than simple DNA transport[268].

1.4.2.4 Proteomic studies of spermatozoa

When we initiated our study in 2001, very little proteomic work had been performed on spermatozoa membrane. As discussed earlier, membrane proteins were hardly amenable to conventional proteomic tools and techniques. One relatively thorough study used cell surface biotinylation. It employed and compared three different extraction buffers for membrane protein enrichment, combined with 2D-gel electrophoresis and protein spot identification by LC-MS/MS[269]. Manual analysis and searching the MS/MS spectra against expressed sequence tag libraries led the authors to conclude that their method had identified at least height novel spermatozoa membrane proteins. More recently, proteomic analysis of whole human spermatozoa using 2D-gel electrophoresis and MALDI-TOF MS allowed the observation of over 1000 different protein species and the identification of 98 proteins, 23% of which had never been observed in this cell previously[270]. In a massive 2D-gel electrophoresis effort to describe human spermatozoa, a combination of overlapping narrow-range pH gradients allowed the establishment of a high-resolution 2D reference map of human spermatozoal proteins from 12 fertile sperm-bank donors containing over 3500 spots, only a few of which have been identified

though[271]. Some recent studies also had additional objectives, other than the simple description of the protein content of the spermatozoa. A group decided to use 2D-DIGE to monitor proteomic changes during the maturation of spermatozoa. Crude cytosolic protein extracts and RP-HPLC of prefractionated cytosolic extracts from spermatogonia, pachytene spermatocytes, and early spermatids were analyzed, revealing over 1200 spots, 245 of which presented significant differences in abundance between the developmental stages[272]. With this approach, 123 proteins were categorized as mitotic, meiotic, and post-meiotic, yielding to the definition of a developmental pattern for normal spermatogenesis. As mentioned earlier, protein phosphorylation is highly implicated in spermatozoa capacitation. To explore this aspect, a phosphoproteomic study was recently performed. The results showed that the majority of the tyrosine kinase activity is present in the cytosolic fraction in bovine spermatozoa, and several tyrosine kinases belonging to three families, Src (Lyn), Csk, and Tec (Bmx, Btk) as well as potential substrates involved in capaciation were identified[273]. Unfortunately, most reported studies are using off-line gel-based approaches which are notorious for being difficult to apply to membrane proteins. To this date, no large-scale gel-free proteomic study of the spermatozoa membrane has been reported.

1.4.2.5 Potential PC4 substrates

As mentioned earlier, PC4 belongs to a family of enzymes responsible for the proteolytic maturation of secretory products. The fact that a PC4 knockout mouse model exhibit impaired fertilization suggests a role for this enzyme and/or its substrates in the fertilization process. Even if the abnormal behaviour of the PC4 deficient spermatozoa was described[226], very little information is available on the enzyme's substrates. The only

confirmed substrate of PC4 in the testes is the precursor of the pituitary adenylate cyclase activating polypeptide (proPACAP). Using RP-HPLC combined with specific radioimmunoassay, it was demonstrated that in the testes of PC4 null mice, proPACAP is not processed to its mature forms, PACAP38 and PACAP27, as it is in wild-type animals[274]. It is hypothesized that this protein could affect sperm function through cAMP mediated signaling, but no other information is available. Recently another endogenous substrate of PC4, pro-insulin-like growth factor-2 (pro-IGF-II) was identified in the human placenta where PC4 was found to be also expressed at very low levels[275]. Considering this very limited knowledge on PC4 and the fact that it presents an unusually restricted expression pattern and a cleavage specificity similar to other members of the PC family, it is quite conceivable that it has one or several other substrates. As the enzyme is present at least at the plasma membrane of spermatozoa[226], a thorough proteomics analysis of the sperm's membrane proteome could allow the identification of other proteins and substrates of PC4 involved in capacitation and/or fertilization. In turn, this could reveal potential therapeutic targets for the treatment of infertility and for the development of new contraceptive strategies.

Our objectives in this study were thus to (i) validate our technological approach using a readily available, though relatively complex, model, (ii) identify new proteins embedded in spermatozoa membrane or intimately associated with it and (iii) assess the presence of PC4 and eventually of putative substrates in the outer face of the membrane.

1.4.3 Dense-core secretory granule

1.4.3.1 The AtT-20 corticotrope cell line

The pituitary gland is a small but complex organ linking the nervous and endocrine systems in the body. The anterior lobe of the pituitary comprises five morphologically distinct endocrine cell types, each producing specific hormones[276]. The nature and effects of the various secretory products of the anterior pituiraty are summarized in Figure 1.11.

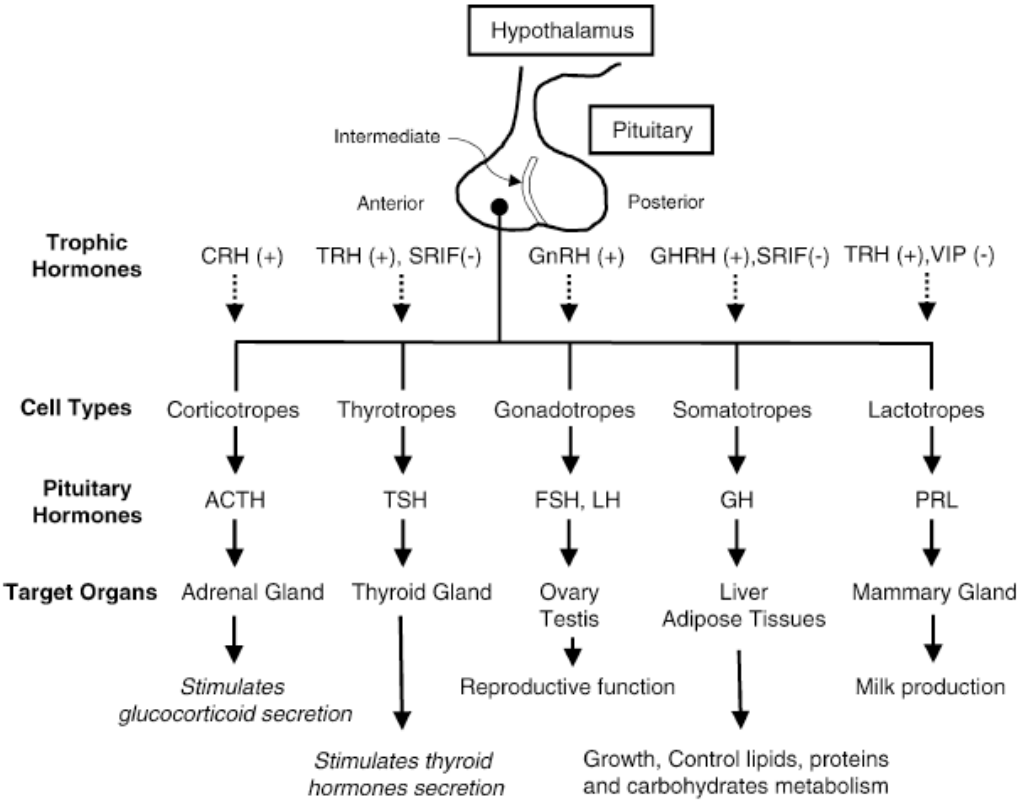


Figure 1.11: Trophic hormones, cell types of the anterior pituitary, their hormonal secretions and their biological effects on target organs[276]. Released by corticotropes in response to stress, ACTH acts on the adrenal glands to induce production and secretion of glucocorticoids.

The cortocotropes represent only approximately 10-20% of all cells in the gland[277]. Since collecting sufficient biological material from dissected glands is very difficult, several lines for each cell type were developed over the years and can be maintained in culture[276]. Reported for the first time in 1975, AtT-20 is an adenocarcinoma-derived, immortalized cell line of murine pituitary corticotropes and a recognized and accepted model for pro-opiomelanocortin (POMC) production and secretion studies[278]. Indeed, as early as 1981, the SG of these cells were isolated using Ficoll gradient centrifugation, studied and shown to contain mature forms adrenocorticotropic hormone (ACTH) and β-lipotropin[279]. Since then, this line of corticotropes has been used in countless studies interested in the production of peptide hormones and their regulated secretion. The major endogenous secretory peptides of AtT-20 cells originate from a single polypeptide precursor, POMC. In cells of the anterior pituitary such as cortocotropes, this precursor undergoes proteolytic maturation under the action of PC1/3 to yield several peptide hormones including ACTH and β-LPH[44]. In melanotropes expressing PC1/3 and PC2, some peptides such as ACTH and βLPH are, for example, further processed into α-MSH and βEnd respectively (Figure 1.12).

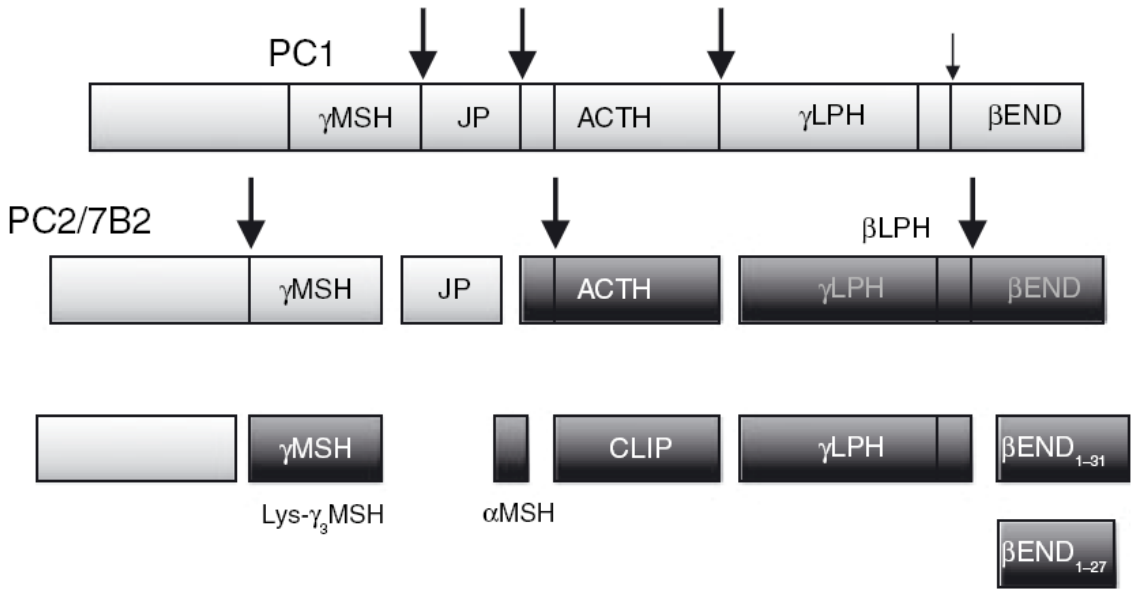


Figure 1.12: Proteolytic processing of proopiomelanocortin and its peptides by prohormone convertases in secretory granules of neuronal and other cells.[44]. In corticotropes, the POMC is mainly converted to ACTH and βLPH by the action of PC1/3. MSH, melanocyte stimulating hormone; JP, joining peptide; ACTH, adrenocorticotropic hormone; LPH, lipotropin; END, endorphin.

1.4.3.2 The regulated pathway of secretion and secretory granule biogenesis

In addition to the constitutive secretory pathway, endocrine and neuroendrocine cells can perform protein exocytosis through a regulated pathway of secretion. Briefly, the genes coding for the proteins destined to this route are translated into mRNA as usual, and their protein products are then co-translationaly inserted into the ER lumen to which they are targeted by virtue of a specific N-terminal sequence known as the signal peptide[280]. Following chaperone-assisted folding and early PTMs, they are exported to the *cis*-face of the Golgi via COPII-coated vesicles[281]. Secretory and membrane proteins will undergo further post-translational modification as they progress through the stacks of the Golgi until they reach the *trans*-Golgi network (TGN). At this point, clathrin-coated immature SGs will bud off the TGN, encompassing their cargo. As they travel along the cytoskeleton toward to cell surface, the SGs will undergo maturation (progressive condensation and acidification), a process during which missorted proteins and clathrin will be removed by the budding off of constitutive like-vesicles. The final, mature SGs containing correctly processed cargo will ultimately reach the cell surface and be ready for secretagogue-induced release of their content. Interestingly, the luminal pH of the secretory pathway plays a critical role in the posttranslational modification and sorting of proteins and lipids. For example, modification and sorting of secretogranin II in PC12 cells[282] or proteolytic maturation of POMC in AtT-20 cells[283] were shown to be dependent on pH. As a consequence, this aspect has been thoroughly investigated using various tools such as pH-sensitive fluorophores[284] and chimeric fluorescent fusion proteins[285]. A generally accepted profile of the pH distribution throughout the secretory pathway is shown in Figure 1.13. As indicated, there is a clear pH gradient along the regulated secretory pathway, from the near-neutral ER to the acidic mature SGs.

Landmarks in the life of the SG are illustrated and described in Figure 1.14. Two crucial processes of the secretory granule's biology will be further discussed herein, namely the sorting and the secretion of the cargo, two events for which several hypotheses and models exist and deserve investigation.

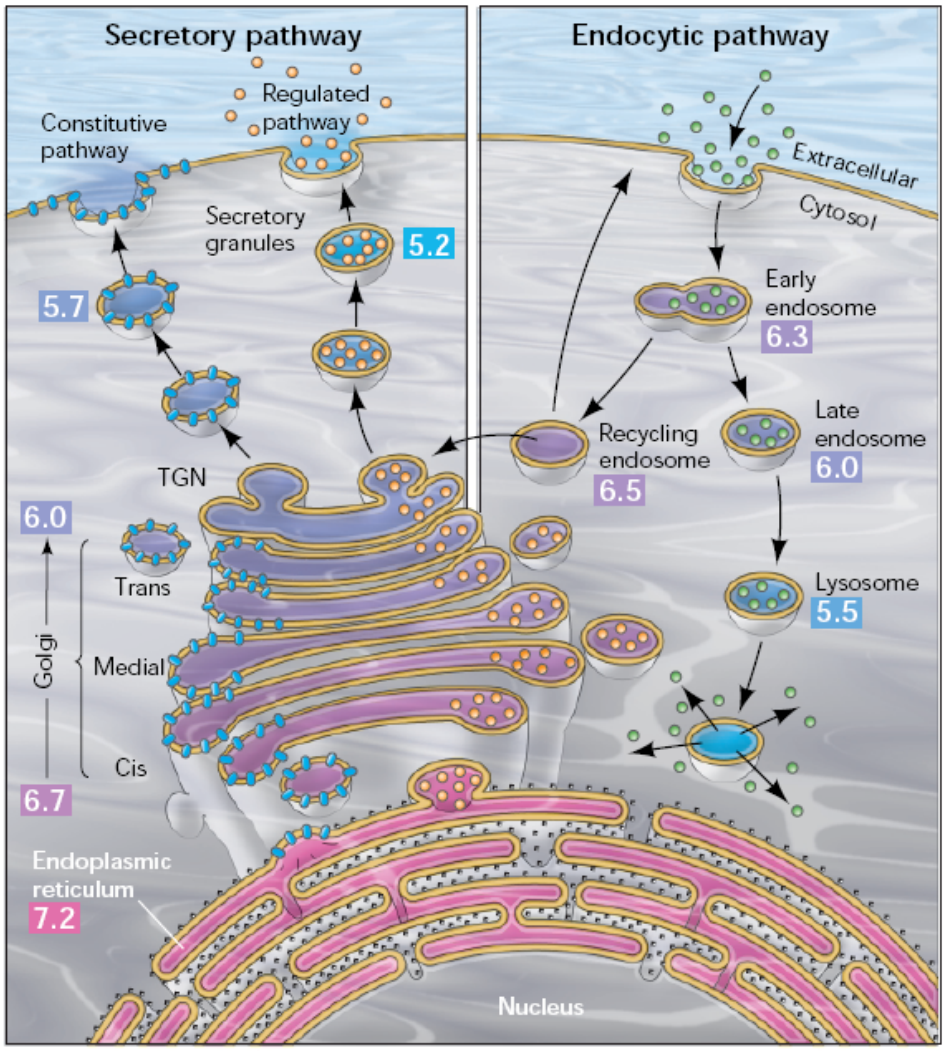


Figure 1.13: Comparison of the steady-state pH of the compartments of the secretory and endocytic pathways[286]. Numbers indicate the approximate luminal pH of the specified compartments. TGN, *trans*-Golgi network

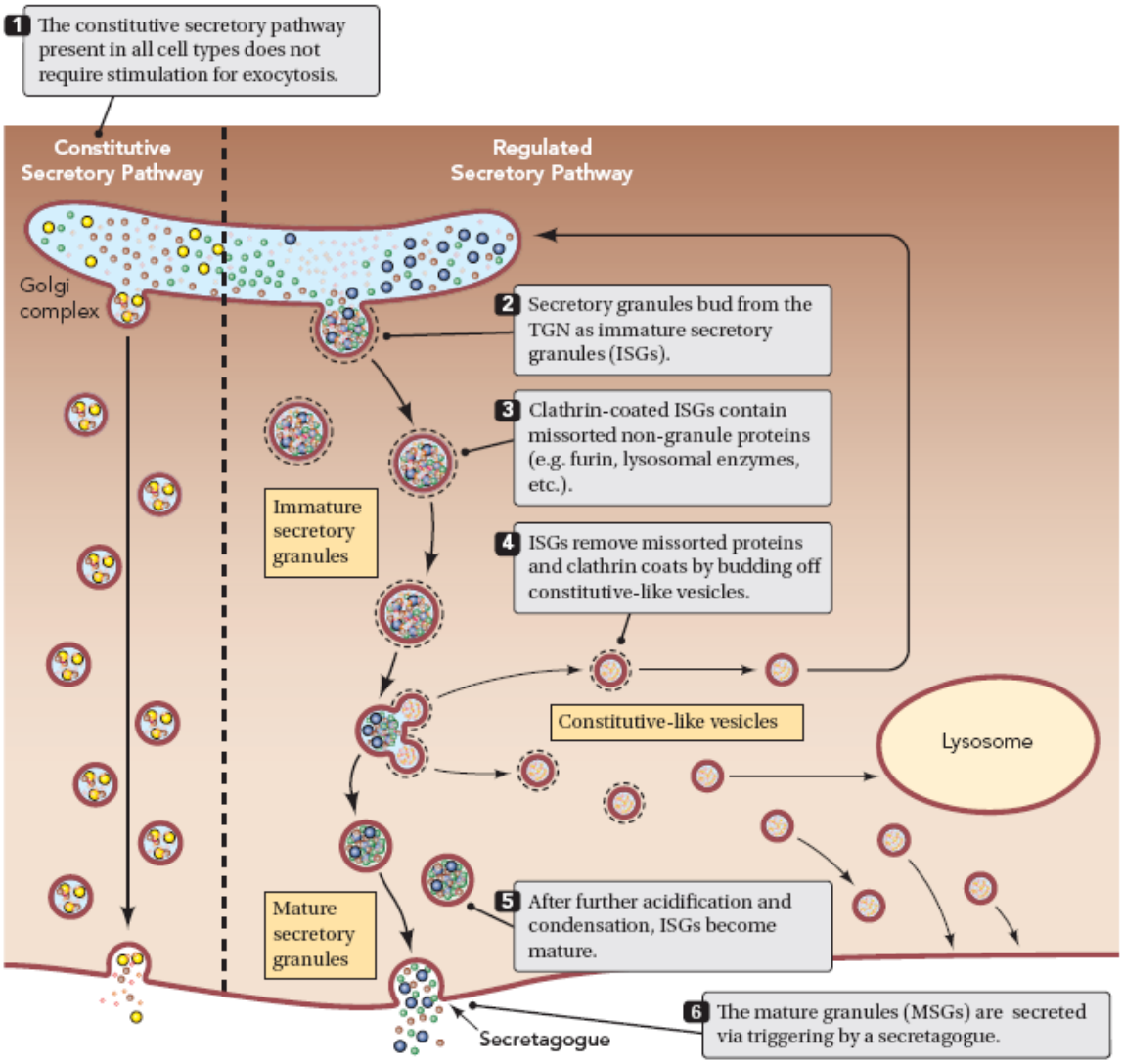


Figure 1.14: Steps involved in secretory granule biogenesis[287]. Two distinctive secretory pathways are present in neuroendocrine and endocrine cells for constitutive and regulated secretion of proteins.

1.4.3.3 Sorting of the cargo in endocrine secretory granules

The processes involved in cargo selection by the SGs have been the subject of many debates, and to date, two hypotheses, the sorting-by-entry and the sorting-by-retention models, are generally considered. These models are not necessarily mutually exclusive[288], and some of the important points of these models as well as the molecules involved will be discussed below.

1.4.3.3.1 Sorting-by-entry

This model proposes that high Ca^{2+} concentration and a pH below 6.5 in specialized parts of the TGN lumen favor the progressive aggregation certain molecules such as the granins. These aggregates would, in turn, act as "seeds" for the precipitation of other regulated cargoes, thereby driving SG formation and maturation[289]. The advantage of the sorting-by-entry model is that it directly couples cargo selection to SG biogenesis. Indeed, it is postulated that granins, including chromogranins A (CgA) and B (CgB) play a granulogenic role by driving budding at the TGN to form SGs. Indeed, the latter and the precursors meant to be targeted to the regulatory pathway of secretion such as provasopressin and POMC have been shown to induce SG formation in fibroblasts and could be considered as important assembly factors[290, 291]. The shared property to aggregate easily in high calcium concentrations and acidic pH could expose at the surface of the proteins and/or aggregates specific areas capable of interacting with membrane proteins or sorting receptors. However, the search for putative "sorting" receptors for cargo proteins has so far only yielded one candidate namely, the CpE enzyme, whose inactivity in $CpE^{fat/fat}$ led to impairments in cargo sorting to the SGs[292] although CpE knock out mice is still viable[293].

1.4.3.3.2 Sorting-by-retention

The second model, sorting-by-retention, could be perceived as complement or refinement of the first. It postulates that not all proteins entering the budding SGs must have been specifically or accurately sorted. However, during SG maturation, missorted proteins would be excluded from the aggregating dense core. They would be removed from the SG along with unnecessary proteins by budding off of clathrin-coated vesicles and re-routed to the constitutive secretory pathway[288]. One of the important pieces of evidence for this model is the case of proinsulin processing by prohormone convertases PC1/3 and PC2[294]. Mature insulin forms insoluble, close packed crystalline arrays in the presence of zinc, whereas proinsulin has distinct physical properties which prevent its condensation into insoluble crystals[295]. Since it has been demonstrated that proinsulin is processed into insulin exclusively inside the SGs[296], not the TGN, and since the sorting-by-entry model depends on aggregation of cargo in the TGN, this model can not adequately explain how proinsulin, which can not aggregate, could still be sorted to the SGs for subsequent processing. More recently, the presence and activity of the clathrin adapator GGA, a Golgi protein, were shown to be essential for SG maturation and removal of the missorted proteins from the immature secretory SGs, emphasizing the importance of this process in SG maturation[297]. As mentioned earlier, a combination of both models can be envisioned to explain sorting of the cargo into SGs and, in both cases, interaction with a receptor would allow selection of the cargo in the forming and maturing SG.

1.4.3.4. Release of the secretory granule content

Regulated delivery of the cargo outside the cell is the ultimate role of the SG. In this case again, several hypotheses and models have been brought forward to explain how this process is accomplished.

In insulin secreting β-cell, the vast majority of SGs are not docked at the membrane, contrary to synaptic vesicles, but appear rather in transit or in close proximity (within 300 nm) to the membrane[298]. They must therefore first be recruited to the cell membrane, a process involving the motor protein kinesin-1, the inactivation of which blocks insulin secretion[299]. However, the recruitment criteria for individual SGs remain largely undefined. Aside from kinesin-1, the only other proteins apparently involved in linking the SG to the microtubule network were identified as kinectin[300] and JIP1-3[301]. Once the SG is within a few hundred nanometer of the cell membrane, movement along the cortical network of microfilaments becomes implicated, and is effected though the action of actin-based motors such as myosin Va[302]. Once more, the other molecular players involved are largely unknown, but a role for the small GTPase Rab27 and of the actin network itself is suggested[303].

Finally, the actual release of the SG's content has been studied for decades, and it is widely accepted that is occurs through exocytosis[304]. This process is mainly accomplished by the interaction of SNAREs with each other, providing both membrane selectivity and the thermodynamic driving force for membrane fusion[303, 305, 306]. In the first model, "full fusion exocytosis", the SG undergoes complete fusion with the plasma membrane and the secretory constituents are entirely discharged from the cell, while the perigranular membrane collapses to and is integrated in the plasma membrane[307].

In the second hypothesis known as the "kiss-and-run" model, a limited surface of the SG and plasma membrane make contact resulting in the opening of smaller and transient pores through which limited and maybe controlled amounts of cargo could be released. Instead of being incorporated to the plasma membrane, the SG's membrane would be immediated retrieved back as was seen using direct observation of fluorescent fusion proteins in endocrine cells[308]. The kiss and run model has been demonstrated using several approaches[304] such as patch-clamp capacitance measurements of the cell surface area, total internal reflectance (TIR)[308] or atomic force microscopes (AFM)[309]. The latter technique was instrumental in identifying a plasma-membrane-associated supramolecular structure called "porosome" where secretory vesicles dock and fuse. Interestingly, the two models, once more, are not mutually exclusive and a review of the most recent and precise electrophysiologcal measurements on single SGs and larger membrane regions clearly indicate that several distinct exocytosis mechanisms occur in pancreatic β-cells[310], a phenomenon that had already been noticed in adrenal chromaffin cells[311] wherein the switch between one mode to the other depends on Ca^{2+} concentrations[312].

Finally, the phenomenon of piecemeal degranulation has gained interest in the last few years. Piecemeal degranulation (PMD) refers to a slow releasing process of stored SG contents mediated by vesicular transport, not membrane fusion[307]. Originally identified in immune cells, the ultrastructural properties characteristic of piecemeal degranulation were observed in several endocrine and neuroendocrine cells, such as adrenal chromaffin cells and pancreatic and α- and β-cells. The growing body of evidence in favor of a contribution of this model to SG secretion was reviewed[307]. Surprisingly, aside from the components of the exocyst complex already responsible for constitutive secretion (Sec6-Sec8)[313] and a few regulators such as small GTPases of the Rab family[314, 315], relatively

little is known on the proteins involved. Understanding how exocytosis occurs for entities such as for the β-cell insulin-containing SG or the corticotrope ACTH-containing SG is of the utmost importance. Not only can it find practical applications in endocrine or neuroendocrine disorders, but it can have more widespread relevance to all closely related organelles[316]. A diagram illustrating the simplified exocytotic-endocytotic cycle of the SG is shown in Figure 1.15.

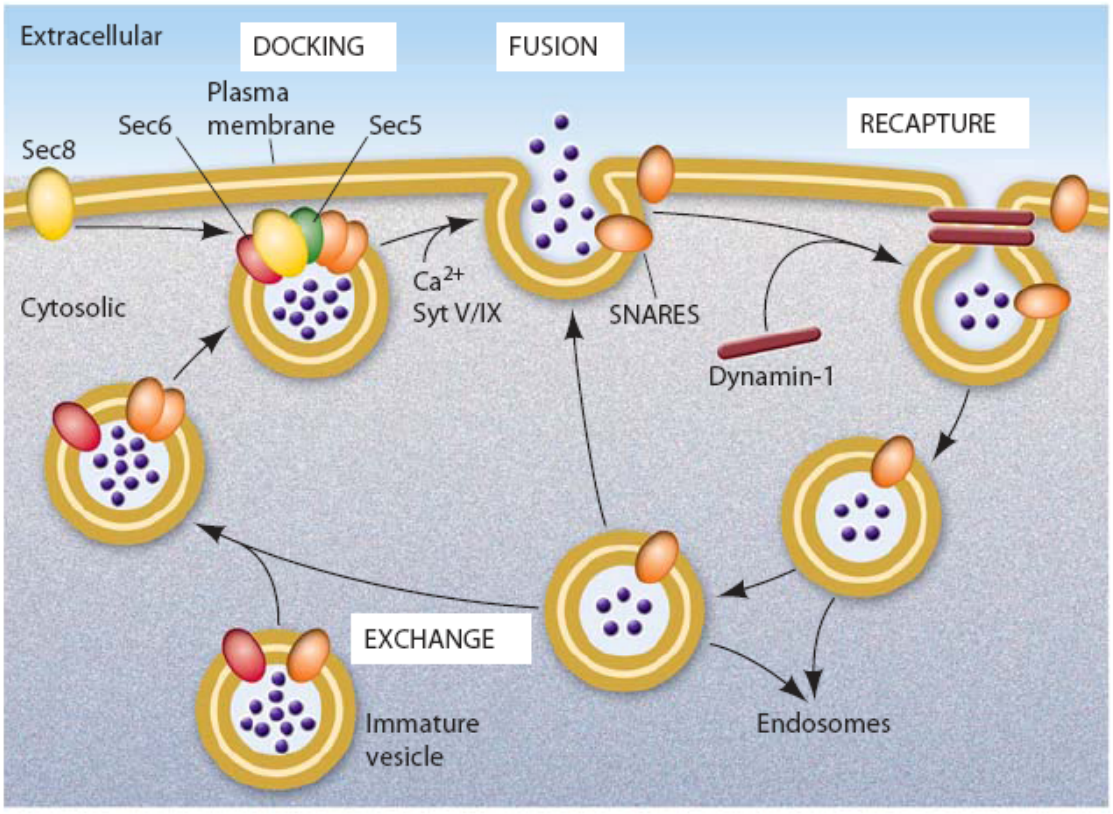


Figure 1.15: Proposed exocytic-endocytic pathways for endocrine secretory granules[303]. Among other players, docking of the SG requires members of the exocyst complex while fusion involves SNAREs. Recruitment of dynamin-1 initiates recapture of the SG which can subsequently fuse with the endosomal network and/or exchange protein material with maturing SGs.

1.4.3.5 Properties and content of the endocrine secretory granule

Dense-core SGs are morphologically distinct, membrane enclosed spherical organelles with diameters of several hundred nanometers[289]. As they bud from the TGN, they are still immature, but the progressive processing and packaging of peptide cargoes lead to the increasing condensation of the SGs' electrodense cores and their conversion into mature SGs[317]. Immunologically, these organelles are also distinguishable by their high immunoreactivity for cell-type-specific peptide hormones and other more widespread cargoes, including CgA, CgB and secretogranins II–VI[318]. Interestingly, mature dense core SGs comes in variable sizes depending upon the organisms, the cell types and even within the same cell[319]. However, as previously mentioned, they are the most acidic compartment of the secretory pathway, with an estimated pH as low as 5.0-5.5[320] and are very rich in Ca^{2+}, with intragranular concentrations as high as 200 mM[321]. Two other divalent cations present, namely $Zn^{2+}$and Cu^{2+} are also relatively specific to the secretory pathway and the SGs[322, 323].

Also, the biology of PC1/3 is directly related to the microenviromental conditions it faces through the secretory pathway. PC1/3 is a calcium-dependent serine protease involved in the processing of prohormones inside the SGs. Like other members of the family, it is synthesized as a zymogen and must first undergo autocatalytic processing before mature PC1/3 can process its substrates[324]. Recombiant PC1/3 was characterized *in vitro* in our laboratory and shown to require calcium and mildly acidic conditions for stability and activity[218]. Since all members of the PC family share nearly identical cleavage site sequence, a challenge in the field is to identify the determinants of the individual convertases' specificity. Since tissular and subcellular localization are not sufficient to explain the differences in substrate specificity, the presence of other factors

and binding partners such as as 7B2[325] and proSAAS[326] could be involved. In summary, several factors endogenous to the SG, ranging from pH to protein content can, as a whole, determine the exact nature of the SG's microenvironment from which several biological processes such as protein aggregation, SG formation[289, 327], peptide hormone maturation or regulated exocytosis will be controlled. Acquiring as much information as possible on this organelle is the only way to elucidate these processes.

Aside from the obvious cell-specific major endocrine product such as insulin for β-cells and POMC for corticotropes, several proteins have been identified as being localized within the dense-core SGs. Among the most abundant are members of the granins family, a group of acidic proteins found in the SGs of endocrine and neuroendocrine cells and composed of CgA[328], CgB[329] and secretogranins (Sg)[330, 331]. Granins tend to bind calcium with low affinity but high capacity and then aggregate *in vitro* at low pH in the presence of calcium, suggesting a role within the core of SGs[318, 332]. They have demonstrated and postulated roles in SG biogenesis[333, 334], regulation of peptide hormone processing[335] and autocrine and paracrine control of secretion[336-338]. Member of this family are very abundant in SGs, and were shown to represent as much as over 80% of all bovine chromaffin granule soluble content[339]. Another interesting class of proteins present in the SG is the prohormone processing enzymes. Among the most important are PC1/3, PC2 and CpE whose role in the proteolytic maturation of peptide hormones from large precursors has already been discussed and reviewed[340-342]. Another prohormone processing enzyme worth of mention is PAM. The peptidylglycine alpha-amidating monooxygenase (PAM) is a bifunctional enzyme that catalyses the conversion of peptidylglycine substrates into alpha-amidated products in a two steps reaction[323], a process that is required for the full activity of nearly half of the bioactive peptides[343]. It

requires copper and ascorbate for activity, and come in several forms, one of which spanning the SG's membrane[323, 344]. A final class of SG proteins is the membrane-associated proteins. In addition to the various coat and SNAREs proteins involved in SG biogenesis, transport and fusion that were already discussed, we must mention the presence of H^+-ATPases proton pumps[345], channels[346] and transporters[347]. Finally, it has long be recognized that there is a significant amount of heterogeneity in the composition of the SGs within a single cell, more so than with other organelles[348, 349]. The biological significance of this phenomenon is still unknown, but certainly deserves further investigation.

1.4.3.6 Proteomic studies of the secretory granules

As described in section 1.3.1.8, the endocrine SG had been neglected as far as proteomic analysis is concerned. Among the numerous reasons for this are its availability, it fragility to experimental manipulations and the difficulty to purify them without contamination from other organelles. Very recently, the analysis of the insulin containing SG[194] and of the chromaffin granule[197] have begun to shed some light on the protein composition of this specialized organelle. Several discrepancies between both studies, both combining density-based centrifugation, 1D-gel electrophoresis and MS-based protein identification, generate questions on the completeness and accuracy of the actual proteomes. Whereas the identification of the abundant granins, some SNAREs and proton pump proteins is clear in both cases, the relatively high abundance of lysosomal hydrolases, especially in the insulin granule, is surprising. Also, the identification of unexpected proteins such as 14-3-3, fructosebisphosphate aldolase A or the peptidyl-prolyl cis-trans isomerase A[194] need to be further confirmed, especially since the scores of

the proteins identified in both studies are sometimes quite low. Clearly, the few proteomic studies already published are not yet sufficient to gain a clear understanding of the endocrine SG's biology. However, even though the secreted cargo differs from one cell type to the other, it is clear that identification of conserved matrix and membrane proteins could provide important insights of the mechanisms by which regulated secretion is accomplished. In turn, those results could help in better understanding misregulation in secretory processes as identified in pathological conditions such as diabetes[310] and neuroendocrine tumors[350].

The objectives of the study of this model, presented in chapter 4, were to i) develop and optimize a new, centrifugation- and gel-free method for the purification and proteomic analysis of corticotropes SGs readily applicable to other organelles, ii) perform fluorescence-assisted purification of the corticotropes mature SGs and define their proteome and iii) identify new SG proteins.

1.5 BIBLIOGRAPHY

1. Alberts, B. *et al.* Molecular biology of the cell. 197 (2002).

2. Dettmer, K., Aronov, P.A. & Hammock, B.D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51-78 (2007).

3. Snoep, J.L. & Westerhoff, H.V. From isolation to integration, a systems biology approach for building the Silicon Cell. *Systems Biology: Definitions and Perspectives*, 425 (2005).

4. Fulton, A.B. & Isaacs, W.B. Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis. *Bioessays* **13**, 157-161 (1991).

5. Wu, S. *et al.* Multi-modality of pI distribution in whole proteome. *Proteomics* **6**, 449-455 (2006).

6. Strachan, R.T., Ferrara, G. & Roth, B.L. Screening the receptorome: an efficient approach for drug discovery and target validation. *Drug Discov. Today* **11**, 708-716 (2006).

7. Anderson, L. Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J. Physiol.* **563**, 23-60 (2005).

8. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

9. Kelleher, N.L. Top-down proteomics. *Anal. Chem.* **76**, 197A-203A (2004).

10. Valaskovic, G.A., Kelleher, N.L. & McLafferty, F.W. Attomole protein characterization by capillary electrophoresis-mass spectrometry. *Science* **273**, 1199-1202 (1996).

11. Anderson, N.G., Matheson, A. & Anderson, N.L. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* **1**, 3-12 (2001).

12. Hu, S., Loo, J.A. & Wong, D.T. Human body fluid proteome analysis. *Proteomics* **6**, 6326-6353 (2006).

13. Pisitkun, T., Shen, R.F. & Knepper, M.A. Identification and proteomic profiling of exosomes in human urine. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13368-13373 (2004).

14. Ball, H.J. & Hunt, N.H. Needle in a haystack: microdissecting the proteome of a tissue. *Amino Acids* **27**, 1-7 (2004).

15. Xu, B.J., Caprioli, R.M., Sanders, M.E. & Jensen, R.A. Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J. Am. Soc. Mass Spectrom.* **13**, 1292-1297 (2002).

16. Emmert-Buck, M.R. *et al.* An approach to proteomic analysis of human tumors. *Mol. Carcinog.* **27**, 158-165 (2000).

17. Pasini, E.M. *et al.* In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* **108**, 791-801 (2006).

18. Kakhniashvili, D.G., Griko, N.B., Bulla, L.A., Jr. & Goodman, S.R. The proteomics of sickle cell disease: profiling of erythrocyte membrane proteins by 2D-DIGE and tandem mass spectrometry. *Exp. Biol. Med. (Maywood)* **230**, 787-792 (2005).

19. Gravel, P. *et al.* Human blood platelet protein map established by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **16**, 1152-1159 (1995).

20. Garcia, A. *et al.* Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* **4**, 656-668 (2004).

21. Thomas, D.P., Pitarch, A., Monteoliva, L., Gil, C. & Lopez-Ribot, J.L. Proteomics to study Candida albicans biology and pathogenicity. *Infect. Disord. Drug Targets* **6**, 335-341 (2006).

22. Kolker, E. *et al.* Initial proteome analysis of model microorganism Haemophilus influenzae strain Rd KW20. *J. Bacteriol.* **185**, 4593-4602 (2003).

23. Carlsohn, E., Nystrom, J., Karlsson, H., Svennerholm, A.M. & Nilsson, C.L. Characterization of the outer membrane protein profile from disease-related Helicobacter pylori isolates by subcellular fractionation and nano-LC FT-ICR MS analysis. *J. Proteome Res.* **5**, 3197-3204 (2006).

24. Govorun, V.M. *et al.* Comparative analysis of proteome maps of Helicobacter pylori clinical isolates. *Biochemistry (Mosc).* **68**, 42-49 (2003).

25. Schmidt, F. *et al.* Complementary analysis of the Mycobacterium tuberculosis proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. *Mol. Cell. Proteomics* **3**, 24-42 (2004).

26. Zhang, C.G., Chromy, B.A. & McCutchen-Maloney, S.L. Host-pathogen interactions: a proteomic view. *Expert Rev. Proteomics* **2**, 187-202 (2005).

27. Biron, D.G., Moura, H., Marche, L., Hughes, A.L. & Thomas, F. Towards a new conceptual approach to "parasitoproteomics". *Trends Parasitol.* **21**, 162-168 (2005).

28. Sims, P.F. & Hyde, J.E. Proteomics of the human malaria parasite Plasmodium falciparum. *Expert Rev. Proteomics* **3**, 87-95 (2006).

29. Makanga, M., Bray, P.G., Horrocks, P. & Ward, S.A. Towards a proteomic definition of CoArtem action in Plasmodium falciparum malaria. *Proteomics* **5**, 1849-1858 (2005).

30. Maxwell, K.L. & Frappier, L. Viral proteomics. *Microbiol. Mol. Biol. Rev.* **71**, 398-411 (2007).

31. Speers, A.E. & Cravatt, B.F. Chemical strategies for activity-based proteomics. *ChemBioChem* **5**, 41-47 (2004).

32. Jessani, N. *et al.* A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods* **2**, 691-697 (2005).

33. Butler, G.S. & Overall, C.M. Proteomic validation of protease drug targets: pharmacoproteomics of matrix metalloproteinase inhibitor drugs using isotope-coded affinity tag labelling and tandem mass spectrometry. *Curr. Pharm. Des.* **13**, 263-270 (2007).

34. Ohtsubo, K. & Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).

35. Kaji, H. *et al.* Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* **21**, 667-672 (2003).

36. Liu, T. *et al.* Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **4**, 2070-2080 (2005).

37. Lee, K.A., Means, G.D. & Patterson, S.D. Phosphoproteomics: Challenges and Opportunities *Curr. Proteomics* **3**, 249-257 (2006).

38. Martin, K. *et al.* Strategies and solid-phase formats for the analysis of protein and peptide phosphorylation employing a novel fluorescent phosphorylation sensor dye. *Comb. Chem. High Throughput Screen* **6**, 331-339 (2003).

39. Sun, X., Chiu, J.F. & He, Q.Y. Application of immobilized metal affinity chromatography in proteomics. *Expert Rev. Proteomics* **2**, 649-657 (2005).

40. Spickett, C.M., Pitt, A.R., Morrice, N. & Kolch, W. Proteomic analysis of phosphorylation, oxidation and nitrosylation in signal transduction. *Biochim. Biophys. Acta* **1764**, 1823-1841 (2006).

41. Finley, D., Ciechanover, A. & Varshavsky, A. Ubiquitin as a central cellular regulator. *Cell* **116**, S29-32, 22 p following S32 (2004).

42. Mayor, T., Lipford, J.R., Graumann, J., Smith, G.T. & Deshaies, R.J. Analysis of polyubiquitin conjugates reveals that the Rpn10 substrate receptor contributes to the turnover of multiple proteasome targets. *Mol. Cell. Proteomics* **4**, 741-751 (2005).

43. Denis, N.J., Vasilescu, J., Lambert, J.P., Smith, J.C. & Figeys, D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. *Proteomics* **7**, 868-874 (2007).

44. Konig, S., Luger, T.A. & Scholzen, T.E. Monitoring neuropeptide-specific proteases: processing of the proopiomelanocortin peptides adrenocorticotropin and alpha-melanocyte-stimulating hormone in the skin. *Exp. Dermatol.* **15**, 751-761 (2006).

45. Petricoin, E.F., Belluco, C., Araujo, R.P. & Liotta, L.A. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat. Rev. Cancer* **6**, 961-967 (2006).

46. Boonen, K. *et al.* Neuropeptides of the islets of Langerhans: A peptidomics study. *Gen. Comp. Endocrinol.* (2007).

47. Fricker, L.D. Neuropeptidomics to study peptide processing in animal models of obesity. *Endocrinology* (2007).

48. Baggerman, G., Cerstiaens, A., De Loof, A. & Schoofs, L. Peptidomics of the larval Drosophila melanogaster central nervous system. *J. Biol. Chem.* **277**, 40368-40374 (2002).

49. Bodzon-Kulakowska, A. *et al.* Methods for samples preparation in proteomic research. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **849**, 1-31 (2007).

50. Ishmukhametov, R.R., Galkin, M.A. & Vik, S.B. Ultrafast purification and reconstitution of His-tagged cysteine-less Escherichia coli F1Fo ATP synthase. *Biochim. Biophys. Acta* **1706**, 110-116 (2005).

51. Southan, C. Has the yo-yo stopped? An assessment of human protein-coding gene number. *Proteomics* **4**, 1712-1726 (2004).

52. Stasyk, T. & Huber, L.A. Zooming in: fractionation strategies in proteomics. *Proteomics* **4**, 3704-3716 (2004).

53. Tannu, N.S. *et al.* Comparative proteomes of the proliferating C(2)C(12) myoblasts and fully differentiated myotubes reveal the complexity of the skeletal muscle differentiation program. *Mol. Cell. Proteomics* **3**, 1065-1082 (2004).

54. De Duve, C., Pressman, B.C., Gianetto, R., Wattiaux, R. & Appelmans, F. Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *Biochem. J.* **60**, 604-617 (1955).

55. Pertoft, H. Fractionation of cells and subcellular particles with Percoll. *J. Biochem. Biophys. Methods* **44**, 1-30 (2000).

56. Holter, H. & Moller, K.M. A substance for aqueous density gradients. *Exp. Cell Res.* **15**, 631-632 (1958).

57. Steck, T.L., Straus, J.H. & Wallach, D.F. A model for the behavior of vesicles in density gradients: implications for fractionation. *Biochim. Biophys. Acta* **203**, 385-393 (1970).

58. Rickwood, D. Centrifugation, a practical approach 2nd edition. (1984).

59. Tannu, N.S. & Hemby, S.E. Methods for proteomics in neuroscience. *Prog. Brain Res.* **158**, 41-82 (2006).

60. Yates, J.R., 3rd, Gilchrist, A., Howell, K.E. & Bergeron, J.J. Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell. Biol.* **6**, 702-714 (2005).

61. Zhao, Y., Zhang, W., Kho, Y. & Zhao, Y. Proteomic analysis of integral plasma membrane proteins. *Anal. Chem.* **76**, 1817-1823 (2004).

62. Howell, K.E., Schmid, R., Ugelstad, J. & Gruenberg, J. Immunoisolation using magnetic solid supports: subcellular fractionation for cell-free functional studies. *Methods Cell Biol.* **31**, 265-292 (1989).

63. Kikuchi, M. *et al.* Proteomic analysis of rat liver peroxisome: presence of peroxisome-specific isozyme of Lon protease. *J. Biol. Chem.* **279**, 421-428 (2004).

64. Lawson, E.L. *et al.* Use of magnetic beads with immobilized monoclonal antibodies for isolation of highly pure plasma membranes. *Electrophoresis* **27**, 2747-2758 (2006).

65. Morciano, M. *et al.* Immunoisolation of two synaptic vesicle pools from synaptosomes: a proteomics analysis. *J. Neurochem.* **95**, 1732-1745 (2005).

66. Wetzel, M.G. & Korn, E.D. Phagocytosis of latex beads by Acahamoeba castellanii (Neff). 3. Isolation of the phagocytic vesicles and their membranes. *J. Cell Biol.* **43**, 90-104 (1969).

67. Stuart, L.M. *et al.* A systems biology analysis of the Drosophila phagosome. *Nature* **445**, 95-101 (2007).

68. Beaufay, H. *et al.* Analytical study of microsomes and isolated subcellular membranes from rat liver. 3. Subfractionation of the microsomal fraction by isopycnic and differential centrifugation in density gradients. *J. Cell Biol.* **61**, 213-231 (1974).

69. Nielsen, P.A. *et al.* Proteomic mapping of brain plasma membrane proteins. *Mol. Cell. Proteomics* **4**, 402-408 (2005).

70. Bernas, T., Gregori, G., Asem, E.K. & Robinson, J.P. Integrating cytomics and proteomics. *Mol. Cell. Proteomics* **5**, 2-13 (2006).

71. Givan, A.L. Flow cytometry: an introduction. *Methods Mol. Biol.* **263**, 1-32 (2004).

72. Murphy, R.F. Analysis and isolation of endocytic vesicles by flow cytometry and sorting: demonstration of three kinetically distinct compartments involved in fluid-phase endocytosis. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8523-8526 (1985).

73. Cossarizza, A., Ceccarelli, D. & Masini, A. Functional heterogeneity of an isolated mitochondrial population revealed by cytofluorometric analysis at the single organelle level. *Exp. Cell Res.* **222**, 84-94 (1996).

74. Dhandayuthapani, S. *et al.* Green fluorescent protein as a marker for gene expression and cell biology of mycobacterial interactions with macrophages. *Mol. Microbiol.* **17**, 901-912 (1995).

75. Bock, G., Steinlein, P. & Huber, L.A. Cell biologists sort things out: Analysis and purification of intracellular organelles by flow cytometry. *Trends Cell Biol.* **7**, 499-503 (1997).

76. Pasquali, C., Fialka, I. & Huber, L.A. Subcellular fractionation, electromigration analysis and mapping of organelles. *J. Chromatogr. B Biomed. Sci. Appl.* **722**, 89-102 (1999).

77. Wilson, R.B. & Murphy, R.F. Flow-cytometric analysis of endocytic compartments. *Methods Cell Biol.* **31**, 293-317 (1989).

78. Aniento, F. & Gruenberg, J. Membrane transport from early to late endosomes. *Cold Spring Harb. Symp. Quant. Biol.* **60**, 205-209 (1995).

79. Nilsson, T., Lucocq, J.M., Mackay, D. & Warren, G. The membrane spanning domain of beta-1,4-galactosyltransferase specifies trans Golgi localization. *EMBO J.* **10**, 3567-3575 (1991).

80. Van Berkel, T.J., Kruijt, J.K. & Koster, J.F. Identity and activities of lysosomal enzymes in parenchymal and non-parenchymal cells from rat liver. *Eur. J. Biochem.* **58**, 145-152 (1975).

81. Guillemin, I., Becker, M., Ociepka, K., Friauf, E. & Nothwang, H.G. A subcellular prefractionation protocol for minute amounts of mammalian cell cultures and tissue. *Proteomics* **5**, 35-45 (2005).

82. Santoni, V., Molloy, M. & Rabilloud, T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070 (2000).

83. Fountoulakis, M. *et al.* Reference map of the low molecular mass proteins of Haemophilus influenzae. *Electrophoresis* **19**, 1819-1827 (1998).

84. Culvenor, J.G. *et al.* Characterization of presenilin complexes from mouse and human brain using Blue Native gel electrophoresis reveals high expression in embryonic brain and minimal change in complex mobility with pathogenic presenilin mutations. *Eur. J. Biochem.* **271**, 375-385 (2004).

85. Chinnasamy, G. & Rampitsch, C. Efficient solubilization buffers for two-dimensional gel electrophoresis of acidic and basic proteins extracted from wheat seeds. *Biochim. Biophys. Acta* **1764**, 641-644 (2006).

86. Rabilloud, T., Adessi, C., Giraudel, A. & Lunardi, J. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**, 307-316 (1997).

87. Herbert, B. Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**, 660-663 (1999).

88. Jiang, L., He, L. & Fountoulakis, M. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J. Chromatogr. A* **1023**, 317-320 (2004).

89. Zuobi-Hasona, K., Crowley, P.J., Hasona, A., Bleiweis, A.S. & Brady, L.J. Solubilization of cellular membrane proteins from Streptococcus mutans for two-dimensional gel electrophoresis. *Electrophoresis* **26**, 1200-1205 (2005).

90. Watkins, L.K. *et al.* Fast C18 solid-phase desalting/delipidation of the human serum apolipoproteins for matrix-assisted laser desorption ionization and electrospray ionization mass spectrometric analysis. *J. Chromatogr. A* **840**, 183-193 (1999).

91. Leimgruber, R.M. *et al.* Development of improved cell lysis, solubilization and imaging approaches for proteomic analyses. *Proteomics* **2**, 135-144 (2002).

92. Gorg, A., Weiss, W. & Dunn, M.J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685 (2004).

93. Lee, J. *et al.* Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. *Mol. Cell. Proteomics* **6**, 669-676 (2007).

94. Kolker, E., Higdon, R. & Hogan, J.M. Protein identification and expression analysis using mass spectrometry. *Trends Microbiol.* **14**, 229-235 (2006).

95. Cleveland, D.W., Fischer, S.G., Kirschner, M.W. & Laemmli, U.K. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* **252**, 1102-1106 (1977).

96. Gorg, A. *et al.* The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-1053 (2000).

97. Scheler, C. *et al.* Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* **19**, 918-927 (1998).

98. Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* **1**, 1852-1858 (2006).

99. Jin, L.T., Hwang, S.Y., Yoo, G.S. & Choi, J.K. A mass spectrometry compatible silver staining method for protein incorporating a new silver sensitizer in sodium dodecyl sulfate-polyacrylamide electrophoresis gels. *Proteomics* **6**, 2334-2337 (2006).

100. White, I.R. *et al.* A statistical comparison of silver and SYPRO Ruby staining for proteomic analysis. *Electrophoresis* **25**, 3048-3054 (2004).

101. Viswanathan, S., Unlu, M. & Minden, J.S. Two-dimensional difference gel electrophoresis. *Nat. Protoc.* **1**, 1351-1358 (2006).

102. Miller, I., Crawford, J. & Gianazza, E. Protein stains for proteomic applications: which, when, why? *Proteomics* **6**, 5385-5408 (2006).

103. Miura, K. Imaging technologies for the detection of multiple stains in proteomics. *Proteomics* **3**, 1097-1108 (2003).

104. Yang, Y., Thannhauser, T.W., Li, L. & Zhang, S. Development of an integrated approach for evaluation of 2-D gel image analysis: Impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow. *Electrophoresis* **28**, 2080-2094 (2007).

105. Frohlich, T. & Arnold, G.J. Proteome research based on modern liquid chromatography--tandem mass spectrometry: separation, identification and quantification. *J. Neural Transm.* **113**, 973-994 (2006).

106. Mitulovic, G. & Mechtler, K. HPLC techniques for proteomics analysis--a short overview of latest developments. *Brief. Funct. Genomic Proteomic* **5**, 249-260 (2006).

107. Shen, Y. *et al.* High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Anal. Chem.* **74**, 4235-4249 (2002).

108. Schaefer, H. *et al.* A peptide preconcentration approach for nano-high-performance liquid chromatography to diminish memory effects. *Proteomics* **4**, 2541-2544 (2004).

109. Link, A.J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676-682 (1999).

110. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242-247 (2001).

111. Issaq, H.J., Chan, K.C., Janini, G.M., Conrads, T.P. & Veenstra, T.D. Multidimensional separation of peptides for effective proteomic analysis. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **817**, 35-47 (2005).

112. Biemann, K. Mass spectrometry of peptides and proteins. *Annu. Rev. Biochem.* **61**, 977-1010 (1992).

113. Edman, P. Sequence determination. *Mol. Biol. Biochem. Biophys.* **8**, 211-255 (1970).

114. Koichi Tanaka, H.W.Y.I.S.A.Y.Y.T.Y.T.M. Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151-153 (1988).

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

116. Wysocki, V.H., Resing, K.A., Zhang, Q. & Cheng, G. Mass spectrometry of peptides and proteins. *Methods* **35**, 211-222 (2005).

117. Hayes, R.N. & Gross, M.L. Collision-induced dissociation. *Methods Enzymol.* **193**, 237-263 (1990).

118. March, R.E. Quadrupole ion trap mass spectrometry: a view at the turn of the century. *Int. J. Mass spectrom.* **200**, 285-312 (2000).

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

120. Yates, J.R., Cociorva, D., Liao, L. & Zabrouskov, V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* **78**, 493-500 (2006).

121. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994-999 (1999).

122. Andersen, J.S. *et al.* Nucleolar proteome dynamics. *Nature* **433**, 77-83 (2005).

123. Chen, X., Sun, L., Yu, Y., Xue, Y. & Yang, P. Amino acid-coded tagging approaches in quantitative proteomics. *Expert Rev. Proteomics* **4**, 25-37 (2007).

124. Wu, W.W., Wang, G., Baek, S.J. & Shen, R.F. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J. Proteome Res.* **5**, 651-658 (2006).

125. Palagi, P.M., Hernandez, P., Walther, D. & Appel, R.D. Proteome informatics I: bioinformatics tools for processing experimental data. *Proteomics* **6**, 5435-5444 (2006).

126. Pappin, D.J., Hojrup, P. & Bleasby, A.J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**, 327-332 (1993).

127. Clauser, K.R., Baker, P. & Burlingame, A.L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871-2882 (1999).

128. Zhang, W. & Chait, B.T. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* **72**, 2482-2489 (2000).

129. Yates, J.R., 3rd, Eng, J.K. & McCormack, A.L. Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* **67**, 3202-3210 (1995).

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

131. Apweiler, R., Bairoch, A. & Wu, C.H. Protein sequence databases. *Curr. Opin. Chem. Biol.* **8**, 76-80 (2004).

132. Nesvizhskii, A.I. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol. Biol.* **367**, 87-119 (2007).

133. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

134. Bader, J.S., Chaudhuri, A., Rothberg, J.M. & Chant, J. Gaining confidence in high-throughput protein interaction networks. *Nat. Biotechnol.* **22**, 78-85 (2004).

135. Hirsh, E. & Sharan, R. Identification of conserved protein complexes based on a model of protein network evolution. *Bioinformatics* **23**, e170-176 (2007).

136. Cusick, M.E., Klitgord, N., Vidal, M. & Hill, D.E. Interactome: gateway into systems biology. *Hum. Mol. Genet.* **14 Spec No. 2**, R171-181 (2005).

137. Tosatto, S.C. & Toppo, S. Large-scale prediction of protein structure and function from sequence. *Curr. Pharm. Des.* **12**, 2067-2086 (2006).

138. Hawkins, T. & Kihara, D. Function prediction of uncharacterized proteins. *J. Bioinform. Comput. Biol.* **5**, 1-30 (2007).

139. Eisenhaber, B. & Eisenhaber, F. Posttranslational modifications and subcellular localization signals: indicators of sequence regions without inherent 3D structure? *Curr. Protein Pept. Sci.* **8**, 197-203 (2007).

140. Gardy, J.L. & Brinkman, F.S. Methods for predicting bacterial protein subcellular localization. *Nat. Rev. Microbiol.* **4**, 741-751 (2006).

141. Betts, M.J. & Russell, R.B. The hard cell: From proteomics to a whole cell model. *FEBS Lett.* (2007).

142. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383-5392 (2002).

143. Sioud, M. Therapeutic siRNAs. *Trends Pharmacol. Sci.* **25**, 22-28 (2004).

144. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).

145. Cheuk, W. & Chan, J.K. Subcellular localization of immunohistochemical signals: knowledge of the ultrastructural or biologic features of the antigens helps predict the signal localization and proper interpretation of immunostains. *Int. J. Surg. Pathol.* **12**, 185-206 (2004).

146. Gruenbaum, Y., Margalit, A., Goldman, R.D., Shumaker, D.K. & Wilson, K.L. The nuclear lamina comes of age. *Nat. Rev. Mol. Cell Biol.* **6**, 21-31 (2005).

147. Wilkie, G.S. & Schirmer, E.C. Guilt by association: the nuclear envelope proteome and disease. *Mol. Cell. Proteomics* **5**, 1865-1875 (2006).

148. Schirmer, E.C. & Gerace, L. Organellar proteomics: the prizes and pitfalls of opening the nuclear envelope. *Genome Biol.* **3**, REVIEWS1008 (2002).

149. Schirmer, E.C., Florens, L., Guan, T., Yates, J.R., 3rd & Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301**, 1380-1382 (2003).

150. Goodchild, R.E., Kim, C.E. & Dauer, W.T. Loss of the dystonia-associated protein torsinA selectively disrupts the neuronal nuclear envelope. *Neuron* **48**, 923-932 (2005).

151. Scherl, A. *et al.* Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell* **13**, 4100-4109 (2002).

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

153. Coute, Y. *et al.* Deciphering the human nucleolar proteome. *Mass Spectrom. Rev.* **25**, 215-234 (2006).

154. Visintin, R. & Amon, A. The nucleolus: the magician's hat for cell cycle tricks. *Curr. Opin. Cell Biol.* **12**, 752 (2000).

155. Horky, M., Kotala, V., Anton, M. & Wesierska-Gadek, J. Nucleolus and apoptosis. *Ann. N. Y. Acad. Sci.* **973**, 258-264 (2002).

156. Sinclair, D.A., Mills, K. & Guarente, L. Molecular mechanisms of yeast aging. *Trends Biochem. Sci.* **23**, 131-134 (1998).

157. Knoblach, B. *et al.* ERp19 and ERp46, new members of the thioredoxin family of endoplasmic reticulum proteins. *Mol. Cell. Proteomics* **2**, 1104-1119 (2003).

158. Gilchrist, A. *et al.* Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265-1281 (2006).

159. Appenzeller-Herzog, C. & Hauri, H.P. The ER-Golgi intermediate compartment (ERGIC): in search of its identity and function. *J. Cell Sci.* **119**, 2173-2183 (2006).

160. Scott, M.S., Thomas, D.Y. & Hallett, M.T. Predicting subcellular localization via protein motif co-occurrence. *Genome Res.* **14**, 1957-1966 (2004).

161. Preisinger, C. *et al.* YSK1 is activated by the Golgi matrix protein GM130 and plays a role in cell migration through its substrate 14-3-3zeta. *J. Cell Biol.* **164**, 1009-1020 (2004).

162. Taylor, R.S., Fialka, I., Jones, S.M., Huber, L.A. & Howell, K.E. Two-dimensional mapping of the endogenous proteins of the rat hepatocyte Golgi complex cleared of proteins in transit. *Electrophoresis* **18**, 2601-2612 (1997).

163. Wu, C.C., Yates, J.R., 3rd, Neville, M.C. & Howell, K.E. Proteomic analysis of two functional states of the Golgi complex in mammary epithelial cells. *Traffic* **1**, 769-782 (2000).

164. Wu, C.C. *et al.* Organellar proteomics reveals Golgi arginine dimethylation. *Mol. Biol. Cell* **15**, 2907-2919 (2004).

165. Rabouille, C. & Klumperman, J. Opinion: The maturing role of COPI vesicles in intra-Golgi transport. *Nat. Rev. Mol. Cell Biol.* **6**, 812-817 (2005).

166. Takatalo, M.S., Kouvonen, P., Corthals, G., Nyman, T.A. & Ronnholm, R.H. Identification of new Golgi complex specific proteins by direct organelle proteomic analysis. *Proteomics* **6**, 3502-3508 (2006).

167. Johnson, D.T. *et al.* Tissue heterogeneity of the mammalian mitochondrial proteome. *Am. J. Physiol. Cell Physiol.* **292**, C689-697 (2007).

168. Prokisch, H. *et al.* Integrative analysis of the mitochondrial proteome in yeast. *PLoS Biol.* **2**, e160 (2004).

169. Mootha, V.K. *et al.* Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629-640 (2003).

170. Taylor, S.W. *et al.* Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol.* **21**, 281-286 (2003).

171. Westermann, B. & Neupert, W. 'Omics' of the mitochondrion. *Nat. Biotechnol.* **21**, 239-240 (2003).

172. Claros, M.G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241**, 779-786 (1996).

173. Higgs, R.E., Knierman, M.D., Gelfanova, V., Butler, J.P. & Hale, J.E. Comprehensive label-free method for the relative quantification of proteins from biological samples. *J. Proteome Res.* **4**, 1442-1450 (2005).

174. Kim, H.K. *et al.* Mitochondrial alterations in human gastric carcinoma cell line. *Am. J. Physiol. Cell Physiol.* (2007).

175. Dencher, N.A., Frenzel, M., Reifschneider, N.H., Sugawa, M. & Krause, F. Proteome alterations in rat mitochondria caused by aging. *Ann. N. Y. Acad. Sci.* **1100**, 291-298 (2007).

176. Balaban, R.S., Nemoto, S. & Finkel, T. Mitochondria, oxidants, and aging. *Cell* **120**, 483-495 (2005).

177. Journet, A., Chapel, A., Kieffer, S., Roux, F. & Garin, J. Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. *Proteomics* **2**, 1026-1040 (2002).

178. Vellodi, A. Lysosomal storage disorders. *Br. J. Haematol.* **128**, 413-431 (2005).

179. Bagshaw, R.D., Mahuran, D.J. & Callahan, J.W. Lysosomal membrane proteomics and biogenesis of lysosomes. *Mol. Neurobiol.* **32**, 27-41 (2005).

180. Blott, E.J. & Griffiths, G.M. Secretory lysosomes. *Nat. Rev. Mol. Cell Biol.* **3**, 122-131 (2002).

181. Casey, T.M., Meade, J.L. & Hewitt, E.W. Organelle proteomics: identification of the exocytic machinery associated with the natural killer cell secretory lysosome. *Mol. Cell. Proteomics* **6**, 767-780 (2007).

182. Haddad, E.K., Wu, X., Hammer, J.A., 3rd & Henkart, P.A. Defective granule exocytosis in Rab27a-deficient lymphocytes from Ashen mice. *J. Cell Biol.* **152**, 835-842 (2001).

183. Pryor, P.R. *et al.* Combinatorial SNARE complexes with VAMP7 or VAMP8 define different late endocytic fusion events. *EMBO Rep.* **5**, 590-595 (2004).

184. Griffiths, G. & Mayorga, L. Phagosome proteomes open the way to a better understanding of phagosome function. *Genome Biol.* **8**, 207 (2007).

185. Garin, J. *et al.* The phagosome proteome: insight into phagosome functions. *J. Cell Biol.* **152**, 165-180 (2001).

186. Desjardins, M. ER-mediated phagocytosis: a new membrane for new functions. *Nat. Rev. Immunol.* **3**, 280-291 (2003).

187. Matthews, L.R. *et al.* Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res.* **11**, 2120-2126 (2001).

188. Takamori, S. *et al.* Molecular anatomy of a trafficking organelle. *Cell* **127**, 831-846 (2006).

189. Coughenour, H.D., Spaulding, R.S. & Thompson, C.M. The synaptic vesicle proteome: a comparative study in membrane protein identification. *Proteomics* **4**, 3141-3155 (2004).

190. Burre, J. *et al.* Analysis of the synaptic vesicle proteome using three gel-based protein separation techniques. *Proteomics* **6**, 6250-6262 (2006).

191. Sudhof, T.C. Synaptic vesicles: an organelle comes of age. *Cell* **127**, 671-673 (2006).

192. Burre, J. & Volknandt, W. The synaptic vesicle proteome. *J. Neurochem.* **101**, 1448-1462 (2007).

193. Hutton, J.C. Insulin secretory granule biogenesis and the proinsulin-processing endopeptidases. *Diabetologia* **37 Suppl 2**, S48-56 (1994).

194. Brunner, Y. *et al.* Proteomic analysis of insulin secretory granules. *Mol. Cell. Proteomics* (2007).

195. Roberts, R. Lysosomal cysteine proteases: structure, function and inhibition of cathepsins. *Drug News Perspect.* **18**, 605-614 (2005).

196. Njus, D., Kelley, P.M. & Harnadek, G.J. The chromaffin vesicle: a model secretory organelle. *Physiologist* **28**, 235-241 (1985).

197. Wegrzyn, J., Lee, J., Neveu, J.M., Lane, W.S. & Hook, V. Proteomics of Neuroendocrine Secretory Vesicles Reveal Distinct Functional Systems for Biosynthesis and Exocytosis of Peptide Hormones and Neurotransmitters. *J. Proteome Res.* **6**, 1652-1665 (2007).

198. Oh, P. *et al.* Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* **429**, 629-635 (2004).

199. Olsen, J.V. *et al.* HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol. Cell. Proteomics* **3**, 82-92 (2004).

200. Olsen, J.V., Nielsen, P.A., Andersen, J.R., Mann, M. & Wisniewski, J.R. Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: mapping of neurotransmitter receptors and ion channels. *Brain Res.* **1134**, 95-106 (2007).

201. Pshezhetsky, A.V. *et al.* Subcellular proteomics of cell differentiation: Quantitative analysis of the plasma membrane proteome of Caco-2 cells. *Proteomics* (2007).

202. Berro, R. *et al.* Identifying the membrane proteome of HIV-1 latently infected cells. *J. Biol. Chem.* **282**, 8207-8218 (2007).

203. Mi, J., Kirchner, E. & Cristobal, S. Quantitative proteomic comparison of mouse peroxisomes from liver and kidney. *Proteomics* **7**, 1916-1928 (2007).

204. Saleem, R.A., Smith, J.J. & Aitchison, J.D. Proteomics of the peroxisome. *Biochim. Biophys. Acta* **1763**, 1541-1551 (2006).

205. Blondeau, F. *et al.* Tandem MS analysis of brain clathrin-coated vesicles reveals their critical involvement in synaptic vesicle recycling. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3833-3838 (2004).

206. Borner, G.H., Harbour, M., Hester, S., Lilley, K.S. & Robinson, M.S. Comparative proteomics of clathrin-coated vesicles. *J. Cell Biol.* **175**, 571-578 (2006).

207. Baginsky, S., Grossmann, J. & Gruissem, W. Proteome analysis of chloroplast mRNA processing and degradation. *J. Proteome Res.* **6**, 809-820 (2007).

208. Ferro, M. *et al.* Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Mol. Cell. Proteomics* **2**, 325-345 (2003).

209. Andaluz, S. *et al.* Proteomic profiles of thylakoid membranes and changes in response to iron deficiency. *Photosynth. Res.* **89**, 141-155 (2006).

210. Jin, J. *et al.* Proteomic, functional, and domain-based analysis of in vivo 14-3-3 binding proteins involved in cytoskeletal regulation and cellular organization. *Curr. Biol.* **14**, 1436-1450 (2004).

211. Meng, X. & Wilkins, J.A. Compositional characterization of the cytoskeleton of NK-like cells. *J. Proteome Res.* **4**, 2081-2087 (2005).

212. Taylor, S.W., Fahy, E. & Ghosh, S.S. Global organellar proteomics. *Trends Biotechnol.* **21**, 82-88 (2003).

213. Vaughn, J.L., Goodwin, R.H., Tompkins, G.J. & McCawley, P. The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). *In Vitro* **13**, 213-217 (1977).

214. TR, A., BR, W., FM, D. & KL, A. The Fall Armyworm: A Bibliography. *Fla. Entomol.* **72** (1989).

215. McCarroll, L. & King, L.A. Stable insect cell cultures for recombinant protein production. *Curr. Opin. Biotechnol.* **8**, 590-594 (1997).

216. Jarvis, D.L. Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production. *Virology* **310**, 1-7 (2003).

217. Marchal, I., Jarvis, D.L., Cacan, R. & Verbert, A. Glycoproteins from insect cells: sialylated or not? *Biol. Chem.* **382**, 151-159 (2001).

218. Boudreault, A. *et al.* Molecular characterization, enzymatic analysis, and purification of murine proprotein convertase-1/3 (PC1/PC3) secreted from recombinant baculovirus-infected insect cells. *Protein Expr. Purif.* **14**, 353-366 (1998).

219. Rabah, N., Gauthier, D.J., Gauthier, D. & Lazure, C. Improved PC1/3 production through recombinant expression in insect cells and larvae. *Protein Expr. Purif.* **37**, 377-384 (2004).

220. Ahmad, M., Sayyed, A.H., Crickmore, N. & Saleem, M.A. Genetics and mechanism of resistance to deltamethrin in a field population of Spodoptera litura (Lepidoptera: Noctuidae). *Pest Manag. Sci.* (2007).

221. Zenner, I., Alvarez, A. & Barreto, S. Influence of parasitism by Chelonus insularis Cresson (Hymenoptera: Braconidae) on the susceptibility of Spodoptera frugiperda (J.E. Smith) (Lepidoptera: Noctuidae) to insecticides. *Neotrop. Entomol.* **35**, 818-822 (2006).

222. Taylor, N.A., Van De Ven, W.J. & Creemers, J.W. Curbing activation: proprotein convertases in homeostasis and pathology. *FASEB J.* **17**, 1215-1227 (2003).

223. Seidah, N.G. *et al.* Testicular expression of PC4 in the rat: molecular diversity of a novel germ cell-specific Kex2/subtilisin-like proprotein convertase. *Mol. Endocrinol.* **6**, 1559-1570 (1992).

224. Mbikay, M. *et al.* Structure of the gene for the testis-specific proprotein convertase 4 and of its alternate messenger RNA isoforms. *Genomics* **20**, 231-237 (1994).

225. Mbikay, M. *et al.* Impaired fertility in mice deficient for the testicular germ-cell protease PC4. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6842-6846 (1997).

226. Gyamera-Acheampong, C. *et al.* Sperm from mice genetically deficient for the PCSK4 proteinase exhibit accelerated capacitation, precocious acrosome reaction, reduced binding to egg zona pellucida, and impaired fertilizing ability. *Biol. Reprod.* **74**, 666-673 (2006).

227. Flesch, F.M. & Gadella, B.M. Dynamics of the mammalian sperm plasma membrane in the process of fertilization. *Biochim. Biophys. Acta* **1469**, 197-235 (2000).

228. Visconti, P.E. *et al.* The molecular basis of sperm capacitation. *J. Androl.* **19**, 242-248 (1998).

229. Abou-Haila, A. & Tulsiani, D.R. Mammalian sperm acrosome: formation, contents, and function. *Arch. Biochem. Biophys.* **379**, 173-182 (2000).

230. Harkema, W., Harrison, R.A., Miller, N.G., Topper, E.K. & Woelders, H. Enhanced binding of zona pellucida proteins to the acrosomal region of the intact boar spermatozoa in response to fertilizing conditions: a flow cytometric study. *Biol. Reprod.* **58**, 421-430 (1998).

231. Visconti, P.E. & Kopf, G.S. Regulation of protein phosphorylation during sperm capacitation. *Biol. Reprod.* **59**, 1-6 (1998).

232. Harrison, R.A. & Gadella, B.M. Bicarbonate-induced membrane processing in sperm capacitation. *Theriogenology* **63**, 342-351 (2005).

233. Darszon, A. *et al.* Calcium channels and Ca2+ fluctuations in sperm physiology. *Int. Rev. Cytol.* **243**, 79-172 (2005).

234. Pukazhenthi, B.S., Wildt, D.E., Ottinger, M.A. & Howard, J. Inhibition of domestic cat spermatozoa acrosome reaction and zona pellucida penetration by tyrosine kinase inhibitors. *Mol. Reprod. Dev.* **49**, 48-57 (1998).

235. Nassar, A. *et al.* Modulation of sperm tail protein tyrosine phosphorylation by pentoxifylline and its correlation with hyperactivated motility. *Fertil. Steril.* **71**, 919-923 (1999).

236. Flesch, F.M., Colenbrander, B., van Golde, L.M. & Gadella, B.M. Capacitation induces tyrosine phosphorylation of proteins in the boar sperm plasma membrane. *Biochem. Biophys. Res. Commun.* **262**, 787-792 (1999).

237. Leclerc, P., de Lamirande, E. & Gagnon, C. Cyclic adenosine 3',5'monophosphate-dependent regulation of protein tyrosine phosphorylation in relation to human sperm capacitation and motility. *Biol. Reprod.* **55**, 684-692 (1996).

238. Benoff, S. Voltage dependent calcium channels in mammalian spermatozoa. *Front. Biosci.* **3**, D1220-1240 (1998).

239. Dorval, V., Dufour, M. & Leclerc, P. Regulation of the phosphotyrosine content of human sperm proteins by intracellular Ca2+: role of Ca2+-adenosine triphosphatases. *Biol. Reprod.* **67**, 1538-1545 (2002).

240. Krasznai, Z. *et al.* Role of the Na+/Ca2+ exchanger in calcium homeostasis and human sperm motility regulation. *Cell Motil. Cytoskeleton* **63**, 66-76 (2006).

241. Rufo, G.A., Jr., Schoff, P.K. & Lardy, H.A. Regulation of calcium content in bovine spermatozoa. *J. Biol. Chem.* **259**, 2547-2552 (1984).

242. Gadella, B.M., Lopes-Cardozo, M., van Golde, L.M., Colenbrander, B. & Gadella, T.W., Jr. Glycolipid migration from the apical to the equatorial subdomains of the sperm head plasma membrane precedes the acrosome reaction. Evidence for a primary capacitation event in boar spermatozoa. *J. Cell Sci.* **108 (Pt 3)**, 935-946 (1995).

243. Sirivaidyapong, S., Bevers, M.M. & Colenbrander, B. Acrosome reaction in dog sperm is induced by a membrane-localized progesterone receptor. *J. Androl.* **20**, 537-544 (1999).

244. Nixon, B. *et al.* The identification of mouse sperm-surface-associated proteins and characterization of their ability to act as decapacitation factors. *Biol. Reprod.* **74**, 275-287 (2006).

245. Topfer-Petersen, E. *et al.* Spermadhesins: a new protein family. Facts, hypotheses and perspectives. *Andrologia* **30**, 217-224 (1998).

246. Manjunath, P., Bergeron, A., Lefebvre, J. & Fan, J. Seminal plasma proteins: functions and interaction with protective agents during semen preservation. *Soc. Reprod. Fertil. Suppl.* **65**, 217-228 (2007).

247. Clark, G.F. & Dell, A. Molecular models for murine sperm-egg binding. *J. Biol. Chem.* **281**, 13853-13856 (2006).

248. Snell, W.J. & White, J.M. The molecules of mammalian fertilization. *Cell* **85**, 629-637 (1996).

249. Miller, D.J., Macek, M.B. & Shur, B.D. Complementarity between sperm surface beta-1,4-galactosyltransferase and egg-coat ZP3 mediates sperm-egg binding. *Nature* **357**, 589-593 (1992).

250. Bleil, J.D. & Wassarman, P.M. Identification of a ZP3-binding protein on acrosome-intact mouse sperm by photoaffinity crosslinking. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 5563-5567 (1990).

251. Leyton, L. & Saling, P. 95 kd sperm proteins bind ZP3 and serve as tyrosine kinase substrates in response to zona binding. *Cell* **57**, 1123-1130 (1989).

252. Wassarman, P.M. Mouse gamete adhesion molecules. *Biol. Reprod.* **46**, 186-191 (1992).

253. Leyton, L. & Saling, P. Evidence that aggregation of mouse sperm receptors by ZP3 triggers the acrosome reaction. *J. Cell Biol.* **108**, 2163-2168 (1989).

254. Arnoult, C., Zeng, Y. & Florman, H.M. ZP3-dependent activation of sperm cation channels regulates acrosomal secretion during mammalian fertilization. *J. Cell Biol.* **134**, 637-645 (1996).

255. Roldan, E.R., Murase, T. & Shi, Q.X. Exocytosis in spermatozoa in response to progesterone and zona pellucida. *Science* **266**, 1578-1581 (1994).

256. Brewis, I.A. *et al.* Tyrosine phosphorylation of a 95 kDa protein and induction of the acrosome reaction in human spermatozoa by recombinant human zona pellucida glycoprotein 3. *Mol. Hum. Reprod.* **4**, 1136-1144 (1998).

257. Lee, M.A., Check, J.H. & Kopf, G.S. A guanine nucleotide-binding regulatory protein in human sperm mediates acrosomal exocytosis induced by the human zona pellucida. *Mol. Reprod. Dev.* **31**, 78-86 (1992).

258. Brandelli, A. Modulation of sperm acrosomal exocytosis by guanyl nucleotides and G-protein-modifier agents. *Biochem. Mol. Biol. Int.* **41**, 1217-1225 (1997).

259. Geng, J.G. *et al.* Expression of a P-selectin ligand in zona pellucida of porcine oocytes and P-selectin on acrosomal membrane of porcine sperm cells. Potential

implications for their involvement in sperm-egg interactions. *J. Cell Biol.* **137**, 743-754 (1997).

260. Mori, E., Baba, T., Iwamatsu, A. & Mori, T. Purification and characterization of a 38-kDa protein, sp38, with zona pellucida-binding property from porcine epididymal sperm. *Biochem. Biophys. Res. Commun.* **196**, 196-202 (1993).

261. Richardson, R.T. & O'Rand, M.G. Site-directed mutagenesis of rabbit proacrosin. Identification of residues involved in zona pellucida binding. *J. Biol. Chem.* **271**, 24069-24074 (1996).

262. Hunnicutt, G.R., Primakoff, P. & Myles, D.G. Sperm surface protein PH-20 is bifunctional: one activity is a hyaluronidase and a second, distinct activity is required in secondary sperm-zona binding. *Biol. Reprod.* **55**, 80-86 (1996).

263. Forsbach, A. & Heinlein, U.A. Intratesticular distribution of cyritestin, a protein involved in gamete interaction. *J. Exp. Biol.* **201**, 861-867 (1998).

264. Primakoff, P., Hyatt, H. & Tredick-Kline, J. Identification and purification of a sperm surface protein with a potential role in sperm-egg membrane fusion. *J. Cell Biol.* **104**, 141-149 (1987).

265. Inoue, N., Ikawa, M., Isotani, A. & Okabe, M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434**, 234-238 (2005).

266. Cuasnicu, P.S., Gonzalez Echeverria, F., Piazza, A.D., Cameo, M.S. & Blaquier, J.A. Antibodies against epididymal glycoproteins block fertilizing ability in rat. *J. Reprod. Fertil.* **72**, 467-471 (1984).

267. Rubinstein, E., Ziyyat, A., Wolf, J.P., Le Naour, F. & Boucheix, C. The molecular players of sperm-egg fusion in mammals. *Semin. Cell Dev. Biol.* **17**, 254-263 (2006).

268. Ainsworth, C. Cell biology: the secret life of sperm. *Nature* **436**, 770-771 (2005).

269. Shetty, J. *et al.* Differential extraction and enrichment of human sperm surface proteins in a proteome: identification of immunocontraceptive candidates. *Electrophoresis* **22**, 3053-3066 (2001).

270. Martinez-Heredia, J., Estanyol, J.M., Ballesca, J.L. & Oliva, R. Proteomic identification of human sperm proteins. *Proteomics* **6**, 4356-4369 (2006).

271. Li, L.W. *et al.* Establishment of a high-resolution 2-D reference map of human spermatozoal proteins from 12 fertile sperm-bank donors. *Asian J. Androl.* **9**, 321-329 (2007).

272. Rolland, A.D. *et al.* Two-dimensional fluorescence difference gel electrophoresis analysis of spermatogenesis in the rat. *J. Proteome Res.* **6**, 683-697 (2007).

273. Lalancette, C., Faure, R.L. & Leclerc, P. Identification of the proteins present in the bull sperm cytosolic fraction enriched in tyrosine kinase activity: a proteomic approach. *Proteomics* **6**, 4523-4540 (2006).

274. Li, M., Mbikay, M. & Arimura, A. Pituitary adenylate cyclase-activating polypeptide precursor is processed solely by prohormone convertase 4 in the gonads. *Endocrinology* **141**, 3723-3730 (2000).

275. Qiu, Q., Basak, A., Mbikay, M., Tsang, B.K. & Gruslin, A. Role of pro-IGF-II processing by proprotein convertase 4 in human placental development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11047-11052 (2005).

276. Ooi, G.T., Tawadros, N. & Escalona, R.M. Pituitary cell lines and their endocrine applications. *Mol. Cell. Endocrinol.* **228**, 1-21 (2004).

277. Horvath, E. & Kovacs, K. Fine structural cytology of the adenohypophysis in rat and man. *J. Electron Microsc. Tech.* **8**, 401-432 (1988).

278. Harrison, R.W., 3rd, Fairfield, S. & Orth, D.N. Evidence for glucocorticoid transport into AtT-20/D-1 cells. *Biochemistry (Mosc).* **14**, 1304-1307 (1975).

279. Gumbiner, B. & Kelly, R.B. Secretory granules of an anterior pituitary cell line, AtT-20, contain only mature forms of corticotropin and beta-lipotropin. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 318-322 (1981).

280. Lingappa, V.R. & Blobel, G. Early events in the biosynthesis of secretory and membrane proteins: the signal hypothesis. *Recent Prog. Horm. Res.* **36**, 451-475 (1980).

281. Gurkan, C., Stagg, S.M., Lapointe, P. & Balch, W.E. The COPII cage: unifying principles of vesicle coat assembly. *Nat. Rev. Mol. Cell Biol.* **7**, 727-738 (2006).

282. Carnell, L. & Moore, H.P. Transport via the regulated secretory pathway in semi-intact PC12 cells: role of intra-cisternal calcium and pH in the transport and sorting of secretogranin II. *J. Cell Biol.* **127**, 693-705 (1994).

283. Schmidt, W.K. & Moore, H.P. Ionic milieu controls the compartment-specific activation of pro-opiomelanocortin processing in AtT-20 cells. *Mol. Biol. Cell* **6**, 1271-1285 (1995).

284. Seksek, O., Biwersi, J. & Verkman, A.S. Direct measurement of trans-Golgi pH in living cells and regulation by second messengers. *J. Biol. Chem.* **270**, 4967-4970 (1995).

285. Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G. & Tsien, R.Y. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6803-6808 (1998).

286. Paroutis, P., Touret, N. & Grinstein, S. The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology (Bethesda)* **19**, 207-215 (2004).

287. Kim, T., Gondre-Lewis, M.C., Arnaoutova, I. & Loh, Y.P. Dense-core secretory granule biogenesis. *Physiology (Bethesda)* **21**, 124-133 (2006).

288. Arvan, P. & Halban, P.A. Sorting ourselves out: seeking consensus on trafficking in the beta-cell. *Traffic* **5**, 53-61 (2004).

289. Borgonovo, B., Ouwendijk, J. & Solimena, M. Biogenesis of secretory granules. *Curr. Opin. Cell Biol.* **18**, 365-370 (2006).

290. Huh, Y.H., Jeon, S.H. & Yoo, S.H. Chromogranin B-induced secretory granule biogenesis: comparison with the similar role of chromogranin A. *J. Biol. Chem.* **278**, 40581-40589 (2003).

291. Beuret, N., Stettler, H., Renold, A., Rutishauser, J. & Spiess, M. Expression of regulated secretory proteins is sufficient to generate granule-like structures in constitutively secreting cells. *J. Biol. Chem.* **279**, 20242-20249 (2004).

292. Cool, D.R. *et al.* Carboxypeptidase E is a regulated secretory pathway sorting receptor: genetic obliteration leads to endocrine disorders in Cpe(fat) mice. *Cell* **88**, 73-83 (1997).

293. Cawley, N.X. *et al.* The carboxypeptidase E knockout mouse exhibits endocrinological and behavioral deficits. *Endocrinology* **145**, 5807-5819 (2004).

294. Halban, P.A. & Irminger, J.C. Sorting and processing of secretory proteins. *Biochem. J.* **299 (Pt 1)**, 1-18 (1994).

295. Tooze, S.A. Biogenesis of secretory granules in the trans-Golgi network of neuroendocrine and endocrine cells. *Biochim. Biophys. Acta* **1404**, 231-244 (1998).

296. Huang, X.F. & Arvan, P. Formation of the insulin-containing secretory granule core occurs within immature beta-granules. *J. Biol. Chem.* **269**, 20838-20844 (1994).

297. Kakhlon, O., Sakya, P., Larijani, B., Watson, R. & Tooze, S.A. GGA function is required for maturation of neuroendocrine secretory granules. *EMBO J.* **25**, 1590-1602 (2006).

298. Olofsson, C.S. *et al.* Fast insulin secretion reflects exocytosis of docked granules in mouse pancreatic B-cells. *Pflugers Arch.* **444**, 43-51 (2002).

299. Varadi, A., Ainscow, E.K., Allan, V.J. & Rutter, G.A. Involvement of conventional kinesin in glucose-stimulated secretory granule movements and exocytosis in clonal pancreatic beta-cells. *J. Cell Sci.* **115**, 4177-4189 (2002).

300. Burkhardt, J.K. In search of membrane receptors for microtubule-based motors - is kinectin a kinesin receptor? *Trends Cell Biol.* **6**, 127-131 (1996).

301. Verhey, K.J. *et al.* Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. *J. Cell Biol.* **152**, 959-970 (2001).

302. Ivarsson, R., Jing, X., Waselle, L., Regazzi, R. & Renstrom, E. Myosin 5a controls insulin granule recruitment during late-phase secretion. *Traffic* **6**, 1027-1035 (2005).

303. Rutter, G.A. & Hill, E.V. Insulin vesicle release: walk, kiss, pause ... then run. *Physiology (Bethesda)* **21**, 189-196 (2006).

304. Burgoyne, R.D. & Morgan, A. Secretory granule exocytosis. *Physiol. Rev.* **83**, 581-632 (2003).

305. Ungar, D. & Hughson, F.M. SNARE protein structure and function. *Annu. Rev. Cell Dev. Biol.* **19**, 493-517 (2003).

306. Stojilkovic, S.S. Ca2+-regulated exocytosis and SNARE function. *Trends Endocrinol. Metab.* **16**, 81-83 (2005).

307. Crivellato, E., Nico, B., Bertelli, E., Nussdorfer, G.G. & Ribatti, D. Dense-core granules in neuroendocrine cells and neurons release their secretory constituents by piecemeal degranulation (review). *Int. J. Mol. Med.* **18**, 1037-1046 (2006).

308. Taraska, J.W., Perrais, D., Ohara-Imaizumi, M., Nagamatsu, S. & Almers, W. Secretory granules are recaptured largely intact after stimulated exocytosis in cultured endocrine cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2070-2075 (2003).

309. Jena, B.P. Cell secretion machinery: studies using the AFM. *Ultramicroscopy* **106**, 663-669 (2006).

310. MacDonald, P.E. & Rorsman, P. The ins and outs of secretion from pancreatic beta-cells: control of single-vesicle exo- and endocytosis. *Physiology (Bethesda)* **22**, 113-121 (2007).

311. Steyer, J.A., Horstmann, H. & Almers, W. Transport, docking and exocytosis of single secretory granules in live chromaffin cells. *Nature* **388**, 474-478 (1997).

312. Ales, E. *et al.* High calcium concentrations shift the mode of exocytosis to the kiss-and-run mechanism. *Nat. Cell. Biol.* **1**, 40-44 (1999).

313. Tsuboi, T. *et al.* Mammalian exocyst complex is required for the docking step of insulin vesicle exocytosis. *J. Biol. Chem.* **280**, 25565-25570 (2005).

314. Kasai, K. *et al.* Rab27a mediates the tight docking of insulin granules onto the plasma membrane during glucose stimulation. *J. Clin. Invest.* **115**, 388-396 (2005).

315. Yaekura, K. *et al.* Insulin secretory deficiency and glucose intolerance in Rab3A null mice. *J. Biol. Chem.* **278**, 9715-9721 (2003).

316. Calakos, N. & Scheller, R.H. Synaptic vesicle biogenesis, docking, and fusion: a molecular description. *Physiol. Rev.* **76**, 1-29 (1996).

317. Glombik, M.M. & Gerdes, H.H. Signal-mediated sorting of neuropeptides and prohormones: secretory granule biogenesis revisited. *Biochimie* **82**, 315-326 (2000).

318. Taupenot, L., Harper, K.L. & O'Connor, D.T. The chromogranin-secretogranin family. *N. Engl. J. Med.* **348**, 1134-1149 (2003).

319. Grabner, C.P., Price, S.D., Lysakowski, A. & Fox, A.P. Mouse chromaffin cells have two populations of dense core vesicles. *J. Neurophysiol.* **94**, 2093-2104 (2005).

320. Urbe, S., Dittie, A.S. & Tooze, S.A. pH-dependent processing of secretogranin II by the endopeptidase PC2 in isolated immature secretory granules. *Biochem. J.* **321 (Pt 1)**, 65-74 (1997).

321. Nicaise, G., Maggio, K., Thirion, S., Horoyan, M. & Keicher, E. The calcium loading of secretory granules. A possible key event in stimulus-secretion coupling. *Biol. Cell.* **75**, 89-99 (1992).

322. Thorlacius-Ussing, O. Zinc in the anterior pituitary of rat: a histochemical and analytical work. *Neuroendocrinology* **45**, 233-242 (1987).

323. Prigge, S.T., Mains, R.E., Eipper, B.A. & Amzel, L.M. New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. *Cell. Mol. Life Sci.* **57**, 1236-1259 (2000).

324. Seidah, N.G. *et al.* The subtilisin/kexin family of precursor convertases. Emphasis on PC1, PC2/7B2, POMC and the novel enzyme SKI-1. *Ann. N. Y. Acad. Sci.* **885**, 57-74 (1999).

325. Mbikay, M., Seidah, N.G. & Chretien, M. Neuroendocrine secretory protein 7B2: structure, expression and functions. *Biochem. J.* **357**, 329-342 (2001).

326. Fortenberry, Y., Hwang, J.R., Apletalina, E.V. & Lindberg, I. Functional characterization of ProSAAS: similarities and differences with 7B2. *J. Biol. Chem.* **277**, 5175-5186 (2002).

327. Rambourg, A., Clermont, Y., Chretien, M. & Olivier, L. Formation of secretory granules in the Golgi apparatus of prolactin cells in the rat pituitary gland: a stereoscopic study. *Anat. Rec.* **232**, 169-179 (1992).

328. Banks, P. & Helle, K. The release of protein from the stimulated adrenal medulla. *Biochem. J.* **97**, 40C-41C (1965).

329. Lee, R.W. & Huttner, W.B. Tyrosine-O-sulfated proteins of PC12 pheochromocytoma cells and their sulfation by a tyrosylprotein sulfotransferase. *J. Biol. Chem.* **258**, 11326-11334 (1983).

330. Fischer-Colbrie, R., Laslop, A. & Kirchmair, R. Secretogranin II: molecular properties, regulation of biosynthesis and processing to the neuropeptide secretoneurin. *Prog. Neurobiol.* **46**, 49-70 (1995).

331. Huttner, W.B., Gerdes, H.H. & Rosa, P. The granin (chromogranin/secretogranin) family. *Trends Biochem. Sci.* **16**, 27-30 (1991).

332. Yoo, S.H. pH- and Ca(2+)-dependent aggregation property of secretory vesicle matrix proteins and the potential role of chromogranins A and B in secretory vesicle biogenesis. *J. Biol. Chem.* **271**, 1558-1565 (1996).

333. Gerdes, H.H. *et al.* The primary structure of human secretogranin II, a widespread tyrosine-sulfated secretory granule protein that exhibits low pH- and calcium-induced aggregation. *J. Biol. Chem.* **264**, 12009-12015 (1989).

334. Kim, T., Tao-Cheng, J.H., Eiden, L.E. & Loh, Y.P. Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis. *Cell* **106**, 499-509 (2001).

335. Westphal, C.H. *et al.* The neuroendocrine protein 7B2 is required for peptide hormone processing in vivo and provides a novel mechanism for pituitary Cushing's disease. *Cell* **96**, 689-700 (1999).

336. Cadman, P.E., Rao, F., Mahata, S.K. & O'Connor, D.T. Studies of the dysglycemic peptide, pancreastatin, using a human forearm model. *Ann. N. Y. Acad. Sci.* **971**, 528-529 (2002).

337. Russell, J., Gee, P., Liu, S.M. & Angeletti, R.H. Inhibition of parathyroid hormone secretion by amino-terminal chromogranin peptides. *Endocrinology* **135**, 337-342 (1994).

338. Wand, G.S., Takiyyuddin, M., O'Connor, D.T. & Levine, M.A. A proposed role for chromogranin A as a glucocorticoid-responsive autocrine inhibitor of proopiomelanocortin secretion. *Endocrinology* **128**, 1345-1351 (1991).

339. Fischer-Colbrie, R., Hagn, C. & Schober, M. Chromogranins A, B, and C: widespread constituents of secretory vesicles. *Ann. N. Y. Acad. Sci.* **493**, 120-134 (1987).

340. Muller, L. & Lindberg, I. The cell biology of the prohormone convertases PC1 and PC2. *Prog. Nucleic Acid Res. Mol. Biol.* **63**, 69-108 (1999).

341. Dhanvantari, S. *et al.* Carboxypeptidase E, a prohormone sorting receptor, is anchored to secretory granules via a C-terminal transmembrane insertion. *Biochemistry (Mosc).* **41**, 52-60 (2002).

342. Billova, S., Galanopoulou, A.S., Seidah, N.G., Qiu, X. & Kumar, U. Immunohistochemical expression and colocalization of somatostatin, carboxypeptidase-E and prohormone convertases 1 and 2 in rat brain. *Neuroscience* **147**, 403-418 (2007).

343. Merkler, D.J. C-terminal amidated peptides: production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzyme Microb. Technol.* **16**, 450-456 (1994).

344. Boudreault, A., Castellucci, V.F., Chretien, M. & Lazure, C. Identification, purification, and characterization of the molecular forms of Aplysia californica peptidylglycine alpha-amidating enzyme. *J. Neurochem.* **66**, 2596-2605 (1996).

345. Taupenot, L., Harper, K.L. & O'Connor, D.T. Role of H+-ATPase-mediated acidification in sorting and release of the regulated secretory protein chromogranin A: evidence for a vesiculogenic function. *J. Biol. Chem.* **280**, 3885-3897 (2005).

346. Geng, X., Li, L., Watkins, S., Robbins, P.D. & Drain, P. The insulin secretory granule is the major site of K(ATP) channels of the endocrine pancreas. *Diabetes* **52**, 767-776 (2003).

347. Essand, M. *et al.* Identification and characterization of a novel splicing variant of vesicular monoamine transporter 1. *J. Mol. Endocrinol.* **35**, 489-501 (2005).

348. Ozawa, H., Picart, R., Barret, A. & Tougard, C. Heterogeneity in the pattern of distribution of the specific hormonal product and secretogranins within the secretory granules of rat prolactin cells. *J. Histochem. Cytochem.* **42**, 1097-1107 (1994).

349. Sobota, J.A., Ferraro, F., Back, N., Eipper, B.A. & Mains, R.E. Not all secretory granules are created equal: Partitioning of soluble content proteins. *Mol. Biol. Cell* **17**, 5038-5052 (2006).

350. Gratzl, M., Breckner, M. & Prinz, C. Mechanisms of storage and exocytosis in neuroendocrine tumors. *Endocr. Pathol.* **15**, 1-16 (2004).

Chapter 2

Utilization of a new biotinylation reagent in the development of a non-discriminatory investigative approach for the study of cell surface proteins.

Daniel J. GAUTHIER[1], Bernard F. GIBBS[2], Nadia RABAH[1] and Claude LAZURE[1].

[1] Institut de recherches cliniques de Montréal, Québec, Canada.

[2] MDS Pharma Services, Montréal, Québec, Canada and McGill University, Montréal, Qc, Canada.

Proteomics, 2004 Dec;4(12): 3783-90[1].

2.1 PREFACE

When we initiated this study in 2000, proteomics still relied heavily on the emerging technology of 2D-gel electrophoresis. However, this method was already notorious for presenting problems with the analysis of hydrophobic membrane-associated proteins[2]. Even with the optimization of the solubilization buffers, extraction of this class of proteins was incomplete and their separation, both by isoelectric focusing and 1D-SDS-PAGE, was not very efficient. Nervertheless, membrane proteins are of great interest because of their involvement in nearly all cellular processes. Their accessibility at the cell surface makes them ideal as therapeutic targets for the pharmaceutical industry. Hence, significant efforts were devoted to develop new methods for the analysis of cell surface proteomes. Tagging these proteins with biotinylation reagents and isolating them by avidin affinity purification had already been performed with relative success[3]. Indeed, the extremely high affinity of the biotin tag of the proteins for the avidin matrix requires the purified proteins to be recovered using very stringent conditions and detergents, hindering therefore further proteomic analytical steps. Thus, the use of "cleavable" biotinylation reagents possessing a disulfide bridge within their linker region sensitive to reduction conditions was explored. However, this approach precluded any further use of the avidin column to which the biotin moiety remained permanently associated. We have therefore decided to circumvent these problems and design one of the few early gel-free approaches for the large scale proteomic study of membrane proteins. To achieve this goal, we have designed and synthesized a novel protein biotinylation reagent, named sulfo-NHS-iminobiotin, and used it to label the surface of cells. Initially designed with the objective of applying it to the identification of spermatozoa membrane-localized endogenous substrates of PC4, we first had to establish a proof-of-concept for this

method. Since Sf9 cells were readily available in the laboratory and the study of their surface proteins was interesting with respect to their involvement in baculovirus-mediated infection, we have first tested our approach on this cell model. Our preliminary results were initially submitted as a brief communication to Proteomics. However, at the request of the reviewers who proved very interested about our innovative approach, we pushed our analysis further and submitted a complete paper. The details of this new approach to study cell surface proteomes as well as the difficulties faced (and usually overcome) are presented in this chapter.

2.2 ABSTRACT

In order to circumvent the various problems encountered during the study of membrane-bound proteins, we designed and synthesized a novel membrane-impermeable biotinylation reagent incorporating chemical properties compatible with this goal. We then developed a non-discriminatory analytical procedure for such studies which overcomes possible selectivity, contamination and solubility problems. The necessary steps (labeling, limited *in situ* proteolysis, affinity purification) are all conducted in mild or near native conditions. This versatile method could provide an accurate picture of the cell surface proteome.

2.3 INTRODUCTION

Membrane proteins are the crucial players involved in the exchange of both information and biological material between a cell and its environment, or an organelle and the cytosolic compartment. Recently, the field of proteomics has been increasingly interested in studying membrane-associated and transmembrane proteins since they

represent potential targets as far as drug development, immunotherapy or cancer biomarkers are concerned. The generally accepted method to retrieve those proteins involves an initial labeling step by which a marker molecule is covalently attached to the proteins of interest and subsequently used for detection and/or affinity purification purposes. In that respect, many commercially available biotinylation reagents are nowadays available to suit this purpose. The biotin-avidin system and its applications have been known for years[4-6]. Biotin, a naturally occurring vitamin involved, for example, in carboxylation reactions during carbohydrate metabolism, displays upon binding to avidin one of the strongest non-covalent associations known in nature[7], with a K_d (dissociation constant) of 10^{-15}. Scientists have taken advantage of this interaction and developed a series of biotin-containing molecules usually comprising three components namely, (i) a chemically reactive group specifically designed to react with a desired chemical function on the target, (ii) a linker region, and (iii) the biotin moiety itself. Therefore, the labeled molecule, be it DNA, protein or peptide, can be separated from a complex mixture using immobilized avidin. For example, we have used this approach in the past to develop a versatile method to detect proteolytic activities[8].

The greatest advantage of this procedure, namely the potency of the biotin-avidin interaction, unfortunately also becomes its main caveat, as elution of the labeled proteins from the avidin support is extremely difficult. Harsh conditions, involving detergents, chaotropes or other disruptive methods which can potentially damage both the ligand and its binding partner, are required to promote dissociation. In the best cases, this recovery process is often incomplete, rendering the procedure hardly reproducible[9, 10]. Alternative compounds have been developed either to facilitate disruption of the complex or to diminish the strength of binding. One strategy has been to introduce a disulfide bridge in

the linker region of the biotinylation reagent in order to allow elution from the avidin support under reducing conditions[11]. However, if employed to label cells or whole tissues, this reagent is highly vulnerable to endogenous, *in vivo* reduction and certain subsets of proteins could loose their label well before they can be retrieved[12]. Furthermore, this elution method leaves the biotin group attached to avidin, consequently preventing reutilization of the support as well as the use of numerous specific biotin detection tools. The use of monomeric avidin, which displays lesser affinity for biotin, is another alternative[13]. However, pre-treatment of the support is required and the significantly reduced affinity can also detrimentally affect reproducibility between samples. Conversely, the use of biotin derivatives such as iminobiotin, displaying a pH-sensitive interaction with avidin or desthiobiotin easily displaced by biotin has also been advocated in numerous studies[14, 15]. Following labeling, extraction of the cells (or isolated membranes) in order to render soluble the labeled proteins is required prior to analysis[16]. The use of a mixture of ionic and non-ionic detergents, often in combination with chaotropic agents, is considered effective in bringing into solution the membrane proteins. However, it is increasingly recognized that this extraction procedure can be selective and incomplete. Indeed, in addition to the numerous ways by which proteins are associated to the membranes (from simple ionic interaction with a partner to complete insertion of many transmembrane regions), there exist also many differences in composition between membrane domains (such as lipid rafts) complicating the extraction[2, 16-18]. Finally, most of the large hydrophobic proteins extracted in this manner will display very low solubility, will tend to aggregate or precipitate and will impede further analytical steps such as 2-D gel electrophoresis[17]. Despite these limitations, recent studies demonstrated that combined biotinylation, purification, and 2-D gel

electrophoresis led to the identification of some proteins that were not expected to be localized at the cell surface, emphasizing the need to study the cell surface proteome[19]. The aim of this paper is to describe a procedure allowing mapping of membrane-bound proteins located on a tissue, on a cell or an organelle at a given time in a given set of conditions. In order to achieve this objective, a novel biotinylation reagent was synthesized with the following properties: water solubility, cell-membrane impermeability and avidin binding properties rendering elution of the retained material specific and done in gentle conditions. Then, using this reagent, we developed a simple and adaptable non-discriminatory analytical method for the study of membrane proteins comprising four steps. Cells are first labeled using a biotinylation reagent in near native conditions, soluble peptides from labeled membrane proteins are generated through *in situ* limited proteolytic digestion of the cell surface, the peptides are isolated via affinity purification and finally recovered in a buffer suitable for HPLC, MALDI-TOF-MS and Q-TOF-MS/MS analyses.

2.4 MATERIALS AND METHODS

2.4.1 Chemical synthesis of sulfo-NHS-iminobiotin

For the synthesis of the sulfo-N-hydroxysuccinimide ester of 2-iminobiotin, we used a procedure adapted from[20]. Briefly, 2-iminobiotin free base (0.8 mmol, 194.6mg) (Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) was dissolved in water by dropwise addition of 5% HCl under gentle heat (40°C). The resulting 2-iminobiotin.HCl salt was recovered by lyophilization. It was dissolved in 3.2 ml N,Ndimethylformamide

under gentle heat, followed by N-hydroxysulfosuccinimide addition (0.8 mmol, 173.7mg) (Molecular Biosciences, Boulder, CO, USA). The resulting solution was brought to 4°C and dicyclohexylcarbodiimide (0.8 mmol, 164.8mg) (Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) was added. This final reaction solution was stirred at room temperature for 16h, and for 3h at 4°C to induce precipitation of the formed dicyclohexylurea. The insoluble dicyclohexylurea was subsequently removed by centrifugation, the supernatant collected and the organic solvent evaporated under vacuum. The residue was recrystallized from hot 2-propanol yielding the final ester in a 56% yield. The structure of the final compound (sulfo-NHS-iminobiotin) was verified and confirmed by NMR using a Varian VXR400S instrument (Varian, Walnut Creek, CA, USA).

2.4.2 Functional characterization of sulfo-NHS-iminobiotin

2.4.2.1 Affinity binding of iminobiotinylated bovine serum albumin.

In order to document the desired characteristics of the reagent, bovine serum albumin (BSA, Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) was diluted in water to a final concentration of 1.0 mg/ml. Sulfo–NHS-iminobiotin was then added to the solution at a concentration of 0.25 mg/ml and the reaction was allowed to proceed at 4°C for 30 minutes. The reaction was stopped by adding Tris base (final concentration of 50 mM) and the solution dialysed against 50 mM ammonium bicarbonate, pH 10.5. Excess amounts of labeled BSA solution (pH 10.5) was then added to Neutravidin coated agarose beads[21] (Pierce Biotechnology Inc., Rockford, IL, USA) and allowed to bind for

30 minutes at room temperature. The beads were washed 2X with 50 mM ammonium bicarbonate pH 10.5, 2X with 50 mM ammonium bicarbonate, 0.5 M NaCl, pH 10.5, and 2X more with 50 mM ammonium bicarbonate pH 10.5. The high salt concentration washes are crucial to ensure complete dissociation of non-specifically bound material. Labeled BSA was eluted using 50 mM ammonium acetate, pH 4.0. Labeling, binding and elution were monitored by SDS-PAGE followed by silver staining or by affinity blotting using HRP-conjugated Neutravidin (Pierce Biotechnology Inc., Rockford, Il, USA). During detection with Neutravidin-HRP, care was taken to keep the pH of all solutions at 10.5, as no signal is observed if the experiment is carried at neutral or acidic pH (data not shown).

2.4.2.2 Localization of sulfo-NHS-iminobiotinylated proteins.

To test whether sulfo-NHS-iminobiotin could be used to label and retrieve fragments of membrane proteins, we first used an affinity cytochemical approach to confirm that it was impermeable to cell membranes and would therefore specifically target cell surface proteins. Briefly, Sf9 (*Spodoptera Frugiperda* ovarian cells) insect cells (Gibco/Invitrogen, Burlington, Ontario, Canada) grown in adhesion in Sf-900 II serum free medium (Gibco/Invitrogen) were washed in ice-cold PBS and fixed in 4% paraformaldehyde for 15 minutes at room temperature. After extensive washings, cells were incubated in a PBS solution containing 0.25 mg/ml sulfo-NHS-iminobiotin or 0.25 mg/ml sulfo-NHS-LC-biotin (Pierce Biotechnology Inc., Rockford, IL, USA) for 30 minutes. The reaction was stopped by adding Tris base to a final concentration of 50 mM and unbound labeling reagents were removed by washing 4X in PBS. Labeled cells were

subsequently incubated for 20 minutes in a PBS solution containing a 1/50 dilution of Alexa Fluor 488-conjugated Streptavidin (Molecular Probes Inc., Eugene, OR, USA) at the indicated pHs and visualized by microscopy. It is noteworthy to mention that Alexa Fluor 488 was chosen since it emits fluorescence in a pH-independent manner.

2.4.3 Affinity purification of Sf9 cell surface peptides

2.4.3.1 Cell surface labeling

For this study, $25x10^{6}$ Sf9 cells were grown in suspension in serum free medium, washed 3X in ice-cold PBS, resuspended in 1.0 ml of ice-cold PBS solution containing 0.25 mg/ml sulfo-NHS-iminobiotin and incubated with gentle agitation for 30 minutes at $4^{\circ}C$. The reaction was stopped by adding Tris base to a final concentration of 50 mM and unbound labeling reagents were removed by washing 5X in PBS.

2.4.3.2 Cell surface limited, *in-situ,* proteolytic digestion

Cells were then submitted to limited *in situ* proteolytic digestion by resuspension in 1.0 ml of a PBS solution containing 15 μg/ml trypsin (Sigma-Aldrich Canada Ltd., Oakville, Ontario, Canada) and incubation for 15 minutes at room temperature with gentle agitation. The enzymatic reaction was stopped by addition of 35 μL of a 1.0 mg/ml TLCK (N-p-Tosyl-lysine chloromethyl ketone, Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) in 1 mM HCl. The washes and partial digests of cell surface proteins were analyzed by SDS-PAGE and affinity blotting.

2.4.3.3 Recovery of cell surface peptides

The resulting labeled peptides were finally purified from the digest using Neutravidin-coated agarose beads, as described above. A saturating amount of labeled Sf9 surface protein digest was analyzed, binding and washes were performed at pH 10.5, and elution achieved at pH 4.0. This procedure was monitored by dot blotting using HRP-conjugated Neutravidin, a method more suited to detect signal originating from all peptides, including those of low molecular weight which are not always visible on gel.

2.4.4 LC-MS/MS analysis and peptide analysis

Protein fragments resulting from cell surface digestion were purified using the above-mentioned procedure and eluted from the Neutravidin-coated bead support using either 50 mM ammonium acetate pH 4.0 or 50% acetonitrile 50% TFA (0.1%). The material contained in the eluates was either directly analyzed by MS or after overnight digestion with trypsin or V8-protease (Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) in ammonium bicarbonate in order to increase the number of peptides and simplify MS/MS sequencing. Samples of crude cell surface digests and of purified membrane protein fragments were then separated and analyzed with a μLC using a Zorbax SB-C_{18} (0.3x250 mm, Phenomenex, Torrance, CA, USA) column coupled to a QSTAR-XL hybrid LC/MS/MS Mass spectrometer (Applied Biosystems, Foster City, CA, USA). The data generated were analyzed using the MASCOT software[22] and the deduced amino acid sequences searched in the complete SWISS-PROT database including all taxa. Finally, the unassigned MS/MS sequences were analyzed individually and more thoroughly by BLAST algorithm searches[23]. Typically, BLAST searches

against arthropods were done using the amino acid sequences determined from the purified samples and the top ranking candidates were evaluated.

2.5 RESULTS AND DISCUSSION

2.5.1 Synthesis and functional characterization of sulfo-NHS-iminobiotin

The structure of the final compound (sulfo-NHS-iminobiotin), as verified by NMR (data not shown), is shown in Figure 2.1 together with other commonly used biotinylation reagents. The resulting compound is water-soluble (by virtue of the addition of the sulfo group) up to a concentration of 0.5 mg/mL which contrasts with the water insolubility of the commercially available NHS-iminobiotin. It is reactive towards primary amines (through the N-hydroxysuccinimidyl group) and incorporates an iminobiotin moiety, which is the cyclic guanidino analog of biotin. The latter has already been shown to bind avidin tightly in its base form (pH>10.0) whereas its protonation in acidic conditions induces a conformational change that is refractory to binding[20, 24]. As illustrated in Figure 2.2, the iminobiotin groups on the labeled BSA can be easily detected by HRP-Neutravidin at basic pH. Upon binding of excess labeled BSA and extensive washing of the beads, elution can be achieved in very mild acidic conditions as seen by both silver staining and affinity blotting. Moreover, no loss of binding capacity was observed following repeated binding and elution experiments conducted on the same beads. These results contrast tremendously with those obtained with any biotin containing labeling reagents requiring harsher conditions. The only other biotinylation reagent considered to be "reversible", namely sulfo-NHS-SS-biotin, does, as previously

mentioned, present some inherent drawbacks. Hence, we conclude that the synthesized product is fully functional and exhibits the required properties.

2.5.2 Affinity purification of sf9 cell surface peptides

To test whether sulfo-NHS-iminobiotin could be used to label and retrieve fragments of membrane proteins, we first used an affinity cytochemical approach to confirm that it was impermeable to cell membranes and would therefore specifically target cell surface proteins. As shown in Figure 2.3, following incubation in the presence of Alexa Fluor 488-conjugated streptavidin at the indicated pH values, only the externally located proteins at the cell surface are observed. The result indicates that sulfo-NHS-iminobiotin labels cells solely at the membrane in a pattern identical to, but at a slightly lesser intensity than sulfo-NHS-LC-biotin, which is cell-membrane impermeable. Collectively, these results confirm that our reagent can reversibly bind an avidin derivative in a pH-dependent manner and provide evidence that it is membrane impermeable and therefore suitable for our study of cell surface proteins.

Labeled cells were then submitted to limited, *in situ*, proteolytic digestion with trypsin and the resulting fragments in washes and from cell digests were analyzed by SDS-PAGE and affinity blotting (Figure 2.4). Results show that during washings, only weak signals are observed at relatively high molecular weights, probably indicating dissociation of loosely bound undigested membrane-associated proteins. However, upon tryptic digestion of the cell surface, an intense signal covering a wide range of molecular weights can be seen, indicative of the liberation of domains and peptides from membrane proteins. The ensuing protein digest was then incubated in the presence of neutravidin coated beads first at pH 10.5 to bind the peptides and, following washes, the peptides

were eluted at pH 4.0. As shown in Figure 2.5, only a very weak signal in the supernatant remains after treatment with the beads. However, an intense signal is recovered upon elution confirming that soluble tryptic fragments of membrane proteins can be recovered and rendered amenable to further analysis by RP-HPLC and mass spectrometric techniques.

2.5.3 MS Analyses

To confirm the effectiveness of our method, we proceeded to the analysis of the samples by mass spectrometry. A representative peptide elution profile and the resulting m/z distribution spectrum from purified Sf9 membrane protein fragments are shown in Figure 2.6. It is noteworthy to mention that despite the single step separation procedure employed, the coupled liquid chromatography Q-TOF analysis led to the direct MS/MS sequencing of over 300 distinct peptides in all samples. Despite the acquisition of several hundreds MS/MS spectra, no significant protein hit could be identified by three or more peptides in the purified samples. This contrasts markedly with the results obtained from the analysis of the unpurified samples in which a large number of Sf9 or insect proteins were rapidly identified using the same search criteria. As an example in a representative unpurified sample, the top 25 significant database hits corresponded to known or putative proteins exhibiting an intracellular localization (ribosomal, cytosolic or nuclear) and furthermore, were assigned to *Spodoptera frugiperda* or closely related insect proteins in a proportion of over 80%. Moreover, none of the peptides assigned to intracellular protein hits were iminobiotinylated, further confirming the membrane impermeability of our labeling reagent. Our results strongly suggest that the purification procedure employed herein is efficient in removing the contribution arising from the presence of intracellular

proteins found in the unpurified samples. This represents a significant result considering the complexity and composition of the starting material. However, in the purified sample which should have contained fragments or peptides derived from the membrane proteins, obtaining convincing identification of Sf9 membrane proteins proved to be very difficult. At this stage, two aspects need to be considered in relation to the techniques and model used, to explain this result. Firstly, the single step separation procedure employed when combined with the restriction of the Mascot public server to 300 peptides analyzed per sample, severely limits our identification capabilities. Secondly, we must emphasize that the relative paucity of membrane protein hits in our samples, despite the large number of MS/MS spectra obtained, is probably due to the absence of such proteins in the available databases. This is further exemplified by the very high and unusual proportion of unassigned peptides in these samples. In addition, the NCBI database lists only 188 reported protein entries for our model organism (Sf9 cells), with only 2.1% of those having a known or a putative primary cell surface localization. This illustrates how under-represented membrane proteins are in the databases, and the need to develop tools to study them. Nevertheless, our analysis of the unpurified sample enabled us to identify a large number of intracellular Sf9 proteins despite the low number of entries in the databases for this organism. Doing so, most database entries displaying identity or very high similarity to the sequences of our purified peptides were uncharacterized *Drosophila melanogaster* or *Anophele gambiae* protein products, often obtained by conceptual translation of the recently available genomic information. Noteworthy, many of the purified peptides in these searches could be assigned to membrane-localized proteins. For example, the peptide sequence RPDDEILK was found to have 8/9 residues identical to *Drosophila* P-glycoprotein 49 (gi266518) an integral membrane protein member of the

multidrug resistance family. Similarly, the sequence TDIPIFK was found to have 5/7 identical and 6/7 chemically similar residues with the exocist component Sec5 (gi24638223), which was recently shown to be an essential controller of the membrane protein trafficking in the *Drosophila* ovary[25]. Finally, the sequence IGKDINTLKTMQEDLIK aligns significantly to three distinct sites of the *Drosophila* lipophorin (gi7511958), an extracellular glycoprotein involved in retinoid and fatty acid binding and transport[26].

2.6 CONCLUDING REMARKS

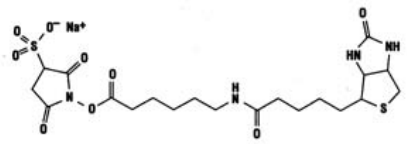
In this paper, we presented two majors improvements in the way cell surface proteins can be studied through proteomics. Firstly, we described the synthesis and functional characterization of a new biotinylation reagent which is water soluble and membrane-impermeable, which allows labeling, binding and elution in mild conditions, and which permits reuse of the avidin support with confidence. Secondly, we used sulfo-NHS-iminobiotin to develop a novel, non-discriminatory investigative approach for membrane proteins. It combines the advantages of (i) labeling in 'near native' conditions, (ii) generating free, soluble fragments of otherwise large and partially hydrophobic proteins and (iii) allowing recovery by affinity retrieval of these peptides in a complete and non-selective manner. Moreover, the peptides are recovered in a buffer that is compatible with further analytical steps, even allowing one to by-pass conventional and 2-D gel electrophoresis which are rather time consuming, have a limited power of resolution and are now perceived as a main bottleneck in most proteomic analyses. One of the expected and observed challenges arising from our approach stems from the

complexity of the iminobiotin-labeled peptides and protein fragments generated. Single step liquid chromatographic separation methods combined with QTOF-MS/MS analysis and computer–assisted database searching using the MASCOT public server were shown to be barely adequate in coping with this aspect. The utilization of refined separation protocols, such as multidimensional separation techniques as developed recently, combined with the precision and sensitivity of QTOF-MS/MS analysis and unrestricted computer–assisted database searching could certainly improve both the quality and the quantity of information generated by a single sample[27, 28]. By requiring low amounts of starting biological material and providing a final sample of highly purified membrane protein soluble fragments, this technique can certainly be of great utility in the study of specific subsets of proteins that still evade, at least in part, conventional proteomics analytical procedures. It is certainly a complement to the cell surface proteome studies employing labeling, extraction and 2-D gel electrophoresis[19]. It is expected that this methodology can find applications in such varied aspects of biological sciences as tissue typing, cell differentiation, tumorigenesis, immunotherapy and the study of pathogenic agents.

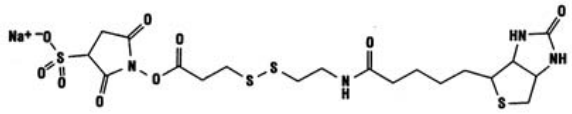
2.7 ACKNOWLEDGMENTS

We would like to thank Dr. Daniel Chapdelaine for NMR characterization and Ms. Dany Gauthier for technical assistance. This study was supported by a Centre of Excellence Protein Engineering Network (PENCE) New Ideas Grant (NIR-7). DJG is a CIHR-funded Ph.D. student.

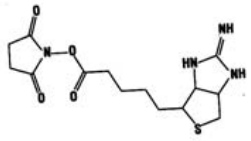
2.8 FIGURES AND LEGENDS



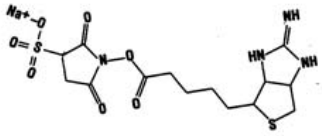
sulfo-NHS-LC-Biotin



sulfo-NHS-SS-Biotin



NHS-Iminobiotin



sulfo-NHS-Iminobiotin

Figure 2.1: The chemical structure of four biotinylation reagents. Sulfo-NHS-LC-biotin, sulfo-NHS-SS-biotin and NHS-iminobiotin are all commercially available whereas we describe herein the synthesis of sulfo-NHS-iminobiotin.

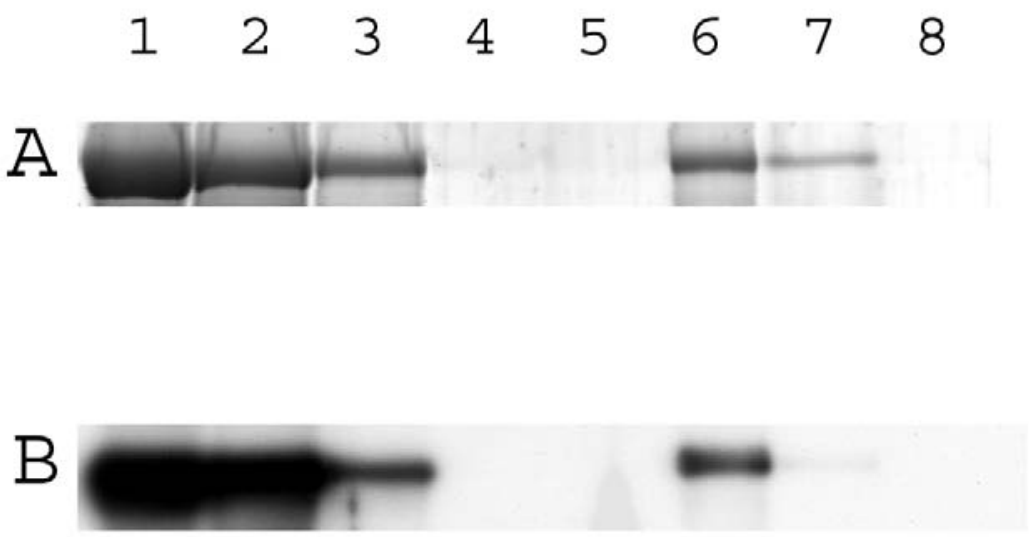


Figure 2.2: Labeling, binding and elution of BSA as described in the text. A) SDS-PAGE followed by silver staining. B) SDS-PAGE followed by affinity blotting. Lane 1: beads supernatant, Lanes 2,3: washes 1 and 2 with 50mM ammonium bicarbonate pH 10.5, Lanes 4,5: following 2 washes with high salt, washes 5 and 6 with 50mM ammonium bicarbonate pH 10.5, Lane 6: eluate 1, Lane 7: eluate 2 and Lane 8: empty.

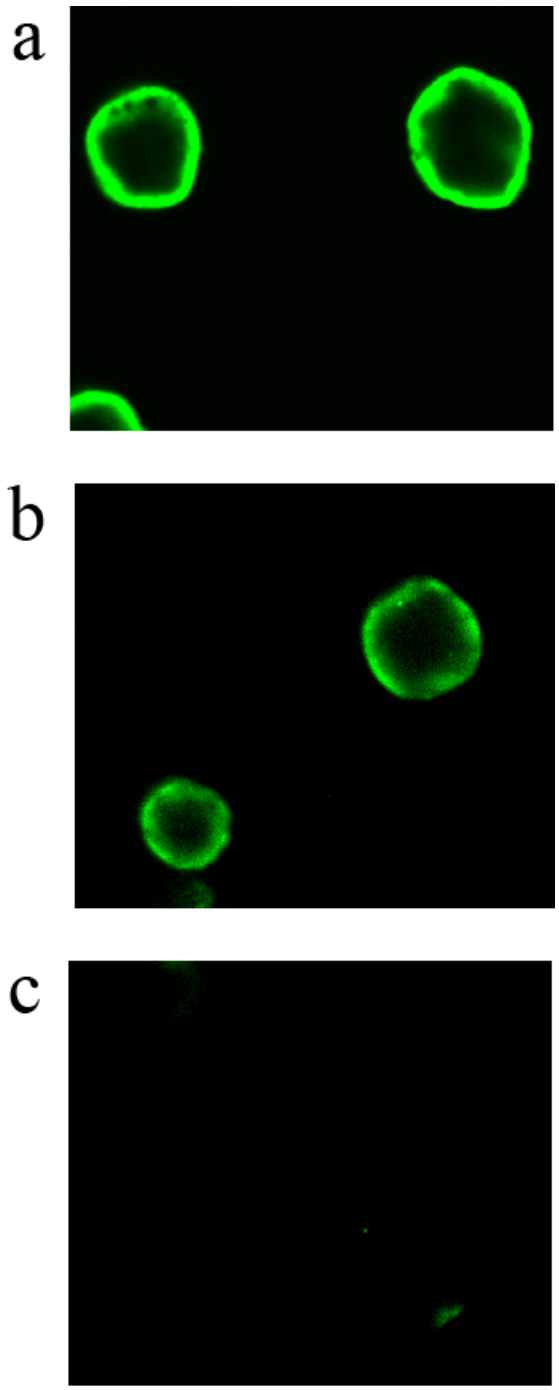


Figure 2.3: Sf9 insect cells were labeled either with sulfo-NHS-LC-biotin (panel A) or sulfo-NHS-iminobiotin (panels B and C). Detection using Alexa Fluor 488-conjugated Streptavidin was done at pH 10.5 (panel A and B) and at pH 4.0 (panel C).

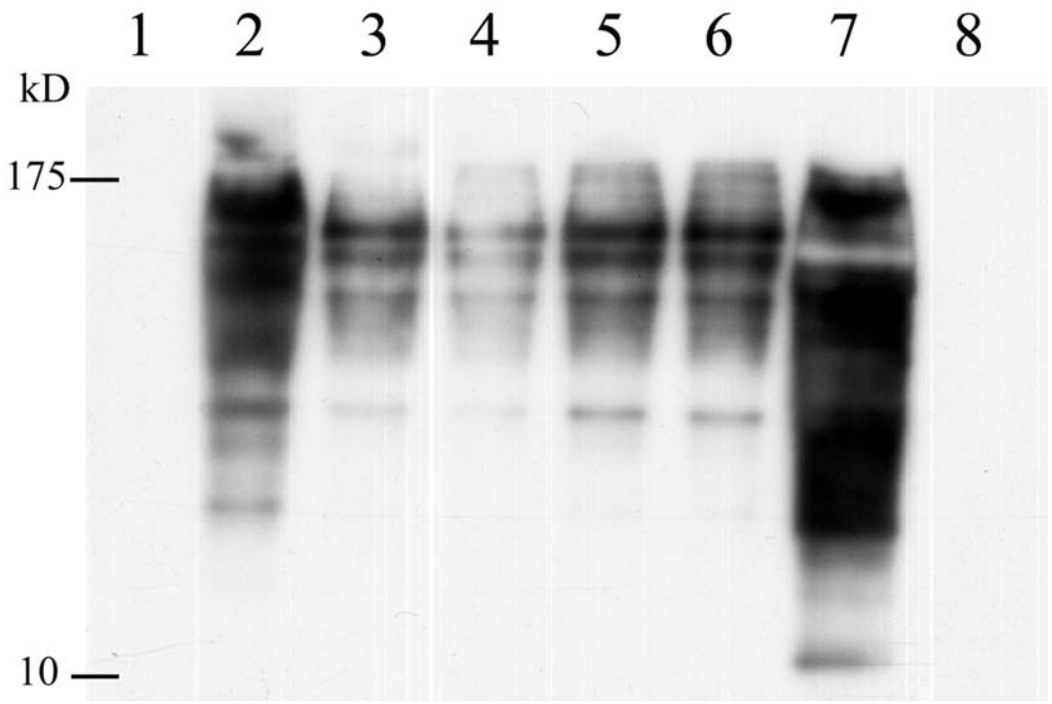


Figure 2.4: Monitoring by affinity blotting of the labeling and limited *in-situ* proteolytic digestion of Sf9 cell surface proteins. Lane 1: Empty, Lanes 2-6: Consecutive washes in ice-cold PBS, Lane 7: Digest with 15 µg/ml trypsin, Lane 8: Empty.

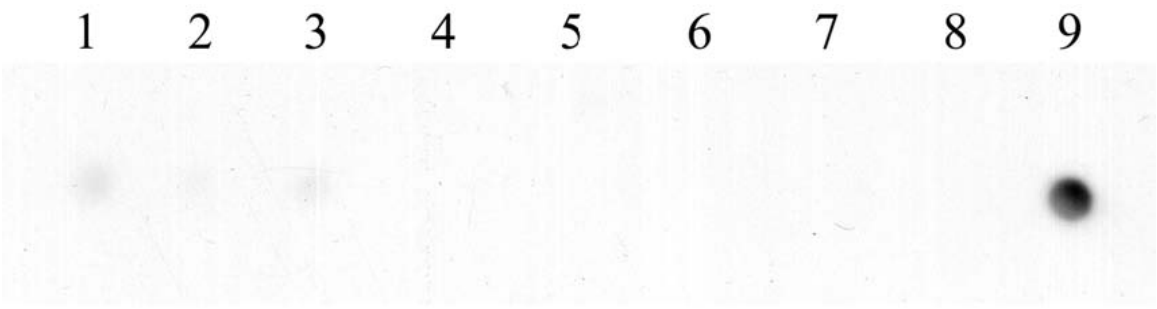


Figure 2.5: Dot blot illustrating affinity purification of labeled membrane protein fragments. Position 1: beads supernatant, Positions 2-3: consecutive washes with 50mM ammonium bicarbonate, pH 10.5, Positions 4-6: washes with 50mM ammonium bicarbonate, 0.5M NaCl, pH 10.5, Positions 7-8: Washes with 50 mM ammonium bicarbonate, pH 10.5, Position 9: pooled eluates in 50mM ammonium acetate, pH 4.0.

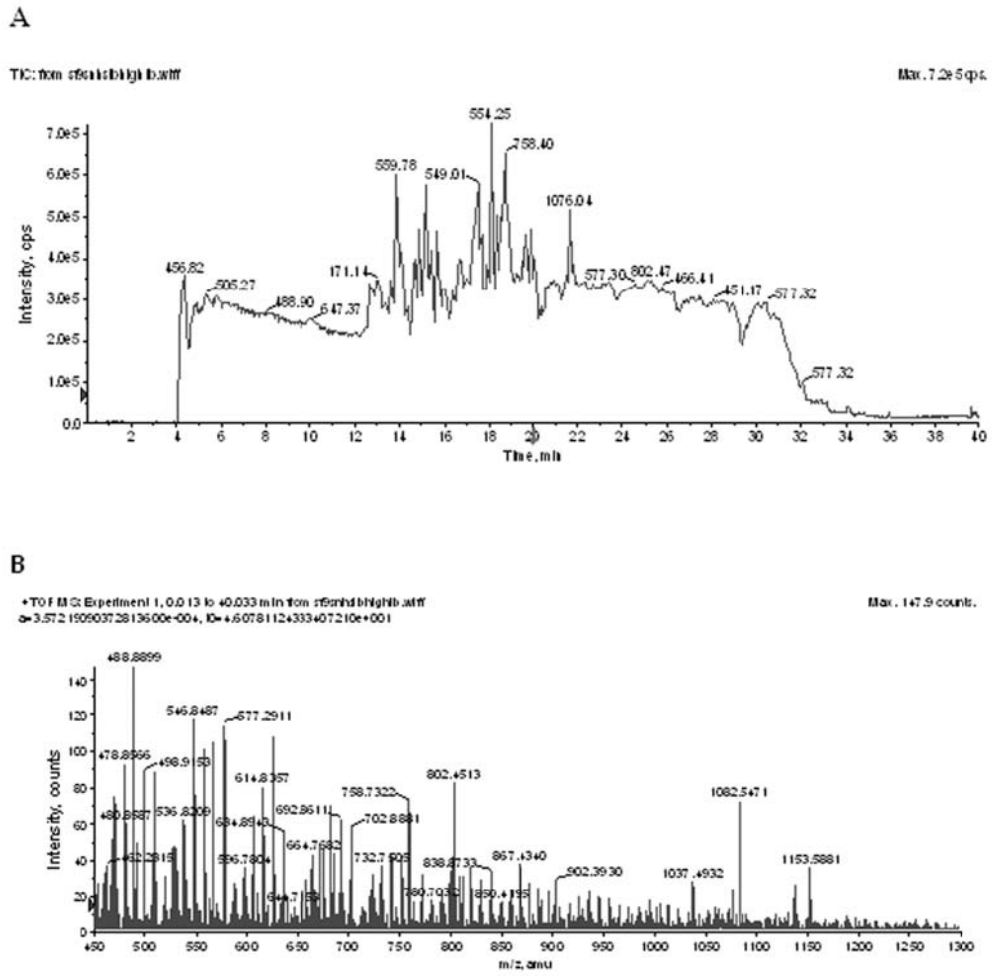


Figure 2.6: Elution profile (A) and ions distribution spectra (B) of purified Sf9 peptides from membrane proteins following separation on a µLC coupled to an QSTAR-MS. Analysis of the sample led to the generation of over 300 distinct MS/MS peptide sequences.

2.9 BIBLIOGRAPHY

1. Gauthier, D.J., Gibbs, B.F., Rabah, N. & Lazure, C. Utilization of a new biotinylation reagent in the development of a nondiscriminatory investigative approach for the study of cell surface proteins. *Proteomics* **4**, 3783-3790 (2004).

2. Santoni, V., Molloy, M. & Rabilloud, T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070 (2000).

3. Kraehenbuhl, J.P. & Bonnard, C. Purification and characterization of membrane proteins. *Methods Enzymol.* **184**, 629-641 (1990).

4. Wilchek, M. & Bayer, E.A. Applications of avidin-biotin technology: literature survey. *Methods Enzymol.* **184**, 14-45 (1990).

5. Bayer, E.A. & Wilchek, M. Application of avidin-biotin technology to affinity-based separations. *J. Chromatogr.* **510**, 3-11 (1990).

6. Diamandis, E.P. & Christopoulos, T.K. The biotin-(strept)avidin system: principles and applications in biotechnology. *Clin. Chem.* **37**, 625-636 (1991).

7. Wilchek, M. & Bayer, E.A. Foreword and introduction to the book (strept)avidin-biotin system. *Biomol. Eng.* **16**, 1-4 (1999).

8. Basak, A., Boudreault, A., Jean, F., Chretien, M. & Lazure, C. Radiolabeled biotinyl peptides as useful reagents for the study of proteolytic enzymes. *Anal. Biochem.* **209**, 306-314 (1993).

9. Cronan, J.E., Jr. Biotination of proteins in vivo. A post-translational modification to label, purify, and study proteins. *J. Biol. Chem.* **265**, 10327-10333 (1990).

10. Gottardi, C.J., Dunbar, L.A. & Caplan, M.J. Biotinylation and assessment of membrane polarity: caveats and methodological concerns. *Am. J. Physiol.* **268**, F285-295 (1995).

11. Shimkus, M., Levy, J. & Herman, T. A chemically cleavable biotinylated nucleotide: usefulness in the recovery of protein-DNA complexes from avidin affinity columns. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2593-2597 (1985).

12. Bickel, U., Kang, Y.S. & Pardridge, W.M. In vivo cleavability of a disulfide-based chimeric opioid peptide in rat brain. *Bioconjug. Chem.* **6**, 211-218 (1995).

13. Bernstein, E.M. & Quick, M.W. Regulation of gamma-aminobutyric acid (GABA) transporters by extracellular GABA. *J. Biol. Chem.* **274**, 889-895 (1999).

14. Garret-Flaudy, F. & Freitag, R. Use of the avidin (imino)biotin system as a general approach to affinity precipitation. *Biotechnol. Bioeng.* **71**, 223-234 (2000).

15. Hirsch, J.D. et al. Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal. Biochem.* **308**, 343-357 (2002).

16. Herbert, B. Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**, 660-663 (1999).

17. Molloy, M.P. Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. *Anal. Biochem.* **280**, 1-10 (2000).

18. Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569-572 (1997).

19. Jang, J.H. & Hanash, S. Profiling of the cell surface proteome. *Proteomics* **3**, 1947-1954 (2003).

20. Orr, G.A. The use of the 2-iminobiotin-avidin interaction for the selective retrieval of labeled plasma membrane components. *J. Biol. Chem.* **256**, 761-766 (1981).

21. Hiller, Y., Gershoni, J.M., Bayer, E.A. & Wilchek, M. Biotin binding to avidin. Oligosaccharide side chain not required for ligand association. *Biochem. J.* **248**, 167-171 (1987).

22. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

23. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

24. Heney, G. & Orr, G.A. The purification of avidin and its derivatives on 2-iminobiotin-6-aminohexyl-Sepharose 4B. *Anal. Biochem.* **114**, 92-96 (1981).

25. Murthy, M. & Schwarz, T.L. The exocyst component Sec5 is required for membrane traffic and polarity in the Drosophila ovary. *Development* **131**, 377-388 (2004).

26. Kutty, R.K. et al. Molecular characterization and developmental expression of a retinoid- and fatty acid-binding glycoprotein from Drosophila. A putative lipophorin. *J. Biol. Chem.* **271**, 20641-20649 (1996).

27. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242-247 (2001).

28. Wu, C.C., MacCoss, M.J., Howell, K.E. & Yates, J.R., 3rd A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21**, 532-538 (2003).

Chapter 3

Determination of the cell surface proteome of freshly ejaculated bovine spermatozoa using a non-discriminatory investigative approach

Daniel J. GAUTHIER[1], Dany GAUTHIER[1], Bernard F. GIBBS[2] and Claude LAZURE[1].

[1] Institut de recherches cliniques de Montréal, Québec, Canada.

[2] MDS Pharma Services, Montréal, Québec, Canada and McGill University, Montréal, Qc, Canada.

Commented draft manuscript from preliminary results.

3.1 PREFACE

Our initial objective while developing an alternative method to study cell surface proteomes was to identify potential endogenous substrates for PC4, a prohormone convertase mainly expressed in male germ cells[1]. In 2002, results already suggested that this enzyme was localized at the surface of mammalian spermatozoa (Gauthier and Lazure, unpublished data) and male mice knock-out for PC4 were infertile[2]. The method presented in chapter 2 showed its potential for the selective enrichment of labeled cell surface protein fragments and their analysis by mass spectrometry. Since most protein players involved in spermatozoa capacitation, ZP binding, acrosome reaction and oocyte binding were already poorly characterized, we decided to apply this newly developed approach to identify the proteins present at the surface of freshly ejaculated bovine spermatozoa. It was reasoned that subsequent analysis of the identified proteins could, in turn, reveal potential endogenous substrates of PC4 involved in the spermatozoa maturation and fertilization processes. Initiated in 2003, this work rapidly yielded very interesting preliminary results, some of which are presented in this chapter. However, difficulties in acquiring large amounts of fresh biological material and restricted availability of mass spectrometry facilities at the time delayed significantly publication of those results. A particular concern on our part was the relatively low score of identified proteins. Very recently, we have been able to bypass the limiting aspects of the study and have resumed our analysis of this cell surface proteome. As mentioned in chapter 5, this work could be expanded to the surface of capacitated and acrosome-reacted spermatozoa and lead to a stage-specific description of the membrane proteins of this cell.

3.2 ABSTRACT

The membrane of the mammalian spermatozoa is a very dynamic structure. Changes, both at the lipid and protein levels are required for freshly ejaculated spermatozoa to acquire the potential to fertilize the oocyte *in vivo*. The sperm cells must undergo a series of sequential maturation steps in order to be able to bind the egg. Spermatozoa are first capacitated during their travel through the female reproductive tract, a process during which there are marked changes in the abundance and distribution of membrane proteins among the different sperm head regions. Once they reach and make contact with the oocyte, spermatozoa will release the content of their acrosome, a vesicle-like structure containing the hydrolytic enzymes necessary for the penetration of the zona pellucida surrounding the egg, and will consequently expose a new membrane surface, a step known as the acrosome reaction. Finally, proteins from the inner acrosomal membrane will interact with their binding partners on the surface of the oocyte which will ultimately lead to fertilization. Despite decades of research, only a few of the membrane-associated proteins involved in the above-mentioned processes have been clearly identified. Interestingly, prohormone convertase 4 (PC4) is a calcium dependent, serine protease involved in the proteolytic maturation of prohormone and proneuropeptide precurors and is almost exclusively present at the surface of spermatozoa. Although very few endogenous substrates for the enzyme have been identified so far, male mice knock-out for PC4 show severely impaired fertility and their sperm present several defects at all stages of maturation, capacitation and fertilization. In order to obtain a clearer picture of the cell-surface exposed proteins of freshly ejaculated bovine spermatozoa, we performed a large-scale, gel-free proteomic analysis of this subcellular proteome and identified over

two hundred proteins, some of which never before reported in this cell type. Further analysis of the identified candidates could reveal their role in the spermatozoa's biology as well as potential substrates for PC4. In turn, this might provide new insights for the development of strategies for the treatment of infertility or of new contraceptive tools.

3.3 INTRODUCTION

The mammalian spermatozoon is a specialized cell comprising several unique structures essential to its function[3]. Interestingly, epididymal and freshly ejaculated spermatozoa do not have the ability to perform fertilization of the oocyte. They can only do so after spending some time in the female reproductive tract where they undergo a series of molecular and physiological transformations collectively known as capacitation[4]. This maturation step characterized, among other things, by changes in the abundance, in the phosphorylation state[5], in the activity[6] and in the localization[7, 8] of several membrane-associated proteins resulting in the increased affinity of the spermatozoa for the zona pellucida surrounding the oocyte[9]. Also, the loss of some cell surface-associated proteins, known as decapacitation factors[10], is also thought to be crucial for this process and to lead to the exposition of new proteins such as the progesterone receptor[11]. Once the sperm make contact with the oocyte, they undergo the acrosome reaction, a process in which the content of their acrosome, a vesicle containing hydrolytic enzymes necessary for the digestion of the zona pellucida and for the penetration of the spermatozoa, is released[12]. Binding of the spermatozoa to the zona pellucida is both carbohydrate- and protein-dependent[13] and is immediately followed by the acrosome reaction. It involves the fusion of the plasma membrane with the outer acrosomal membrane resulting in the ultimate

exposition of a new structure, the inner acrosomal membrane, at the surface of the sperm[14]. A few of the newly exposed membrane proteins involved in the secondary binding have been better characterized. Among the most interesting candidates implicated, we denote P-selectin[15], sp38[16], acrosin[17] and PH-20[18]. Finally, binding to the oocyte plasma membrane is another step in which membrane proteins are very important. Some of the important proteins involved in this binding are cyritestin[19] and fertilin (PH-30)[20] two membrane proteins containing a desintegrin and metalloprotease domain, which are therefore members of the ADAM family.

Prohormone convertase 4 (PC4) is a member of a family of calcium dependent serine proteases. Collectively, these enzymes are involved in the proteolytic maturation of a wide range of secreted and surface-expressed proteins such as hormones, neuropeptides and receptors[21, 22]. Interestingly, the expression pattern of PC4 is almost exclusively restricted to male germ cells, where at least one of the forms of the protein is present at the surface of spermatozoa[23]. Male mice in which the gene coding for PC4 has been inactivated show severely impaired fertility, and their spermatozoa themselves exhibit reduced fertilizing ability and accelerated capacitation, precocious acrosome reaction and reduced binding to the zona pellucida[2, 23]. Contrary to other prohormone convertases, the endogenous substrates of PC4 are mostly unknown. The only one identified so far is proPACAP, the precursor of the pituitary adenylate cyclase activating polypeptide, whose processing is impaired in PC4 knock-out mice in the testis[24]. However, since PACAP is thought to act mainly via an intracrine mechanism inside spermatids and spermatozoa[25], it is reasonable to expect that other substrates of PC4 could exist at the surface of the cell to explain the multiple defects observed in PC4 knock-out spermatozoa.

Clearly, the membrane of the spermatozoa is a very dynamic structure in which changes are indicative of the fertilization potential. Moreover, relatively few proteins have nowadays been conclusively identified as playing a role in the capacitation and fertilization processes. Acquiring the cell surface proteome of these cells could serve two important purposes. First, it would allow identification of all proteins present at every specific maturation step consequently revealing candidates essential to their completion. Secondly, it would permit the identification of potential endogenous substrates for PC4 required for reproductive function. As far as proteomic analysis of the spermatozoa is concerned, several studies have recently been published. Using whole cell extracts, 2D-gel electrophoresis and MALDI-TOF mass spectrometry, 98 human proteins were identified and 1000 distinct protein spots were observed[26]. In another extensive 2D-gel electrophoresis effort to describe human spermatozoa, a combination of overlapping narrow-range pH gradients allowed the presentation of an high-resolution 2D reference map of over 3500 spots representing human spermatozoal proteins, though only a few have been identified[27]. Finally, another study concentrated its efforts on the head region of murine epidydimal spermatozoa. Using surface biotinylation, affinity purification, gel electrophoresis and MS/MS, 85 distinct proteins specific to the membrane were identified[28]. At this point however, the number of identified proteins is relatively low and no specific effort targeting freshly ejaculated bovine spermatozoa has been reported. Moreover, all studies employ detergent extraction and gel electrophoresis which is not very efficient to obtain and resolve certain protein populations and unavoidably lead to sample loss.

In this study, we describe the utilization of a non-discriminatory gel-free proteomic method to purify and analyze proteins associated to the membrane of freshly

ejaculated bovine spermatozoa. Using this approach, we have identified over 200 proteins, several of which had not been reported in this model prior to the present study. Further analysis of the results could reveal important candidates involved in capacitation and fertilization, some of which possibly as endogenous substrates for the prohormone convertase 4.

3. 4 MATERIALS AND METHODS

3.4.1 Bovine spermatozoa acquisition and processing

A fresh bovine ejaculate was obtained from the Centre d'Insémination Artificielle du Québec (CIAQ) and maintained at 18 degrees during transportation to the laboratory (1h). The semen was diluted with 3 volumes PBS and centrifuged at room temperature at 250 x g for 5 minutes. The pellet was washed three times with PBS to remove all traces of seminal plasma and unbound proteins. The cells were finally diluted to a concentration of $1x10^9$ spermatozoa per mL in cold PBS.

3.4.2 Spermatozoa extraction for 1D- and 2D-gel electrophoresis and Western Blot

For 1D gel electrophoresis and Western Blot analyzes of prohormone convertase 4 (PC4), $250x10^6$ spermatozoa were extracted sequentially in 1 mL of two different buffers. The gentle extraction was performed by simply incubating the cell in PBS at 37°C for 15 minutes and collecting the supernatant (extract 1). Subsequently, the cell were resuspended in PBS and submitted to 30 strokes of pestle homogenization, and the supernatant was collected (extract 2). Finally, the cells were resuspended in PBS containing 1% NP-40 (Roche Diagnostics, Laval, QC, Canada) and submitted to 30

strokes of pestle homogenization (extract 3). The same volume of each extract was diluted with SDS-PAGE sample buffer and separated by conventional 10% SDS-PAGE electrophoresis[29]. Proteins were electrotransfered onto an Immobilon-P membrane (Millipore, Billerica, MA, USA) for probing by Western blot. The various PC4-like immunoreactive forms were revealed by a rabbit anti-PC4 antibody (Alexis Biochemicals, Burlington, ON, Canada) and an anti-rabbit secondary antibody coupled to HRP (Chemicon, Temecula, CA, USA) using enhanced chemiluminescence (GE Healthcare, Baie d'Urphée, QC, Canada).

For 2D-gel electrophoresis and Western blotting, spermatozoa were extracted using reagent 2 of the Readyprep sequential kit (8 M urea, 4% CHAPS, 40 mM Tris, 0.2% Bio-Lyte 3/10 ampholyte) (Biorad, Hercules, CA, USA). 25 µg of protein sample was loaded by active rehydration into 7 cm IPG strips for several pH ranges. Isoelectric focusing was performed as per manufacturer's instructions on a Protean IEF cell (Biorad) for a total of approximately 10000 Volt-Hours. Following equilibration as per manufacturer's instructions in SDS-PAGE equilibration buffers I and II, each strip was applied onto a 10% SDS-PAGE gel. Following separation, the proteins were transferred for Western blotting and the presence of PC4-like immunoreactive forms revealed as described above.

3.4.3 Spermatozoa labeling, limited cell surface proteolytic digestion and peptide purification

For the detergent-free, gel free proteomic analysis of the spermatozoa's membrane associated proteins, cells were treated according to a protocol previously described[30]. Briefly, $250x10^6$ cells were labeled for 30 minutes at 4°C in PBS containing 0.5 mg/mL

sulfo-NHS-iminobiotin. The reaction was quenched with 50 mM final Tris, cells were washed and the surface proteins were digested by incubation in 1 mL PBS containing 100 µg/mL trypsin (Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) for 15 minutes. The digestion was stopped by addition of 50 µL of a 1.0 mg/ml TLCK (N-p-Tosyl-lysine chloromethyl ketone, Sigma-Aldrich Canada Ltd, Oakville, Ontario, Canada) in 1 mM HCl. The released, labeled peptides from membrane-associated proteins were purified by affinity using a column of neutravidin-coated agarose beads (Pierce, Rockford, IL, USA). Briefly, 5 mL of beads mixture was washed three times in PBS and incubated for 1 hour in 0.3 mg/mL DMP (Dimethyl pimelimidate Pierce), a protein cross-linking agent, in order to prevent neutravidin shedding from the column during elution. The beads were then used to pack a 10-mL column. The supernatant of the digested spermatozoa solution was brought to pH 10.8 with concentrated NaOH and supplemented with 250 mM final NaCl, applied to the column and left to incubate for 30 minutes at room temperature. The column was then washed with 3 CV of solution W1 (25 mM ammonium bicarbonate, 250 mM NaCl in Chromasolv® water (Sigma-Aldrich) pH 10.8), 3 CV of solution W2 (25 mM ammonium biocarbonate, 500 mM NaCl pH 10.8) and finally 3 CV of solution W1. The bound peptides were eluted with 3 CV of solution E (10% acetonitrile, 25 mM ammonium acetate, acetic acid pH 2.5). The content of each elution fractions was monitored by Dot blotting using HRP-conjugated neutravidin (Pierce) and ECL and the most enriched fractions were pooled and dried under *vacuo*.

3.4.4 LC-MS analysis and protein identification

The material contained in the pooled, dried eluates was analyzed after overnight extensive digestion with trypsin in 25 mM ammonium bicarbonate in order to increase the

number of peptides and simplify MS/MS sequencing. The purified membrane protein fragments were then separated and analyzed with a μLC system using a Zorbax SB-C_{18} (0.3x250 mm, Phenomenex, Torrance, CA, USA) column coupled to a QSTAR-XL hybrid LC/MS/MS Mass spectrometer (Applied Biosystems, Foster City, CA, USA). The data generated was analyzed using the MASCOT software[31] on the MatrixScience public server and the deduced amino acid sequences searched in the complete SWISS-PROT database concentrating on mammalian taxa. Finally, the unassigned proteins and MS/MS sequences were analyzed individually and more thoroughly by BLAST algorithm searches[32]. The quality of the MS/MS spectra was manually verified. Typically, BLAST searches against mammalian sequences were done using the amino acid sequences determined from the purified samples and the top ranking candidates were evaluated. Search results from eight sample were combined to maximize the number of proteins identified.

3. 5 RESULTS AND DISCUSSION

3.5.1 Characterization of the various PC4 forms

As mentioned in the introduction, characterization of the PC4 gene and mRNA revealed that alternative splicing could possibly yield numerous 3' and 5' variants of the mRNA[33]. The biological significance of this splicing in the development of male germ cells is still elusive. However, the identification of an intracellular substrate for the enzyme[24] and the confirmation of the membrane localization of PC4[23] clearly suggest that several protein isoforms could exist in spermatozoa, either sequentially or simultaneously. In order to investigate this aspect, we have performed sequential extractions of freshly

ejaculated bovine spermatozoa in conditions of increasing stringency but in buffers that do not induce significant levels of capacitation or acrosome reaction. The results, presented in Figure 3.1, demonstrate that at least 3 protein products are recognized by the anti-PC4 antibody, at approximately 45 kDa, 54 kDa, 56 kDa. Interestingly, the 54 kDa form is very easily shed from intact cells by simple incubation in PBS at 37°C for 15 minutes. Moreover, this shedding is complete since additional incubation in PBS and/or mechanical stress in this buffer do not result in further 54kDa protein release. Also, protein assays done on this initial extract reveal very insignificant levels of proteins, confirming the microscopic observation in favor of the cells being mostly intact and not having undergone lysis or acrosome reaction in this buffer (data not shown). It is conceivable that this release of a PC4 form could be due to a specific proteolytic activity or to the cleavage of a GPI-anchor from a membrane-bound precursor. The final extract contains a membrane-disrupting detergent and reveals the presence of the additional 45kDa and 56kDa forms. The nature of all three immunoreactive species is not determined yet, and whether they originate from alternative mRNAs or from proteolytic processing of a single polypeptide, as it is the case for other members of the PC family, must be further investigated. However, this data combined with the recently published immunoelectron microscopy results[23] clearly demonstrate that at least one form of the enzyme is present at the surface of intact spermatozoa. Also, narrow-range 2D-gel electrophoresis patterns for these PC4 forms reveal that both the 56 kDa and 45 kDa forms present post-translational modifications affecting the charge and the mass of the proteins, as indicated by a series of adjacent spots (Figure 3.2). The diversity in the pI of the protein forms also might suggest that they could be optimally active in different microenvironments. In summary, early molecular characterization of PC4 show the

potential for multiple forms of the enzyme, at least one of which is present at the surface of the spermatozoa. Considering the infertility phenotype of PC4 knock-out mice, it is possible that this enzyme could have multiple endogenous substrates, some of them present at the cell surface and involved in capacitation and fertilization.

3.5.2 Purification of labeled membrane protein fragments

Sulfo-NHS-iminobiotin synthesized in our laboratory was previously characterized and shown to be membrane-impermeable and to efficiently and specifically label cell surface proteins[30]. Among its main advantages with respect to other biotinylation reagents, it displays a pH-dependent, fully reversible binding to an avidin support. We monitored the binding and elution of the spermatozoa surface protein fragments by dot blotting. As shown in Figure 3.3, initial binding to the neutravidin column at pH 10.8 is near-complete, and the several subsequent washes (low and high salt concentrations, pH 10.8) do not cause bound material to elute. The numerous washes and the utilization of neutravidin, which minimizes non-specific binding, are sufficient to remove the majority if not all non-labeled material. The use of a volatile acidic buffer (Figure 3.3, row D) leads to rapid desorption of the bound material while minimizing possible interferences during subsequent LC-MS/MS analysis.

3.5.3 LC-MS/MS analysis and protein identification

LC-MS/MS analysis from eight independent samples was combined and led to the identification of 203 unique proteins from 321 protein hits. In cases where the MASCOT search results were ambiguous or pointing to an unknown sequence, we have used the BLAST algorithm for short, near exact sequences and the peptide sequence from the

MS/MS data. Usually this procedure led to quick and easy identification. Nearly 65% of all proteins identified have been reported as being permanently or transiently exposed at the cell surface in the literature, confirming the validity of our procedure to enrich and identify membrane-associated proteins. Moreover, over 100 proteins identified had never been reported before in this cell model, the mammalian spermatozoa. A representative list of interesting proteins identified and their proposed or recognized role in spermatozoa biology is presented in Table 3.1. The proteins were also categorized as shown in the scheme depicted in Figure 3.4. A remarkable proportion of the reported proteins can function as membrane receptors or are involved in signal transduction. Another large proportion is directly involved in the dynamics of cellular structure or is part of the glycocalyx (extracellular matrix). Moreover, some unexpected proteins, such as subunits of the proteasome and/or members of the ryanodine receptor family were conclusively identified in this study. Never unambiguously reported in 2003 at the time when these results were obtained, they have since been confirmed to be present at the surface of spermatozoa[34, 35]. The full list of all identified proteins from 8 distinct samples is presented in Supplementary Table 3.1

3.6 CONCLUDING REMARKS

The work presented herein convincingly demonstrates the validity of our detergent-free, gel-free approach to study cell surface proteins, thereby circumventing the problems associated with the solubilization and the separation of this hydrophobic subset of proteins. Applying this approach to the study of the proteins exposed at the surface of freshly ejaculated bovine spermatozoa has allowed the identification of over 200 proteins.

Among these, most of them have been reported to be associated with the plasma membrane and/or were involved in several cellular functions such as signaling, maintenance of cell architecture and membrane fusion. Further analysis of the identified proteins will be necessary to discriminate the candidates involved in capacitation and fertilization, as well as those potentially acting as endogenous susbstrates for PC4. Only then can promising therapeutic targets for contraception or the treatment of infertility be identified.

3.7 ACKNOWLEDGMENTS

We would like to thank the Centre d'insémination artificelle du Québec (CIAQ) for generous semen donation. This study was supported by a Centre of Excellence Protein Engineering Network (PENCE) New Ideas Grant (NIR-7). DJG is a CIHR-funded Ph.D. student.

3.8 CLOSING REMARKS

As previously mentioned, due to technical limitations in 2003, this project was set aside temporarily. The protein samples generated were very pure and fully compatible with mass spectrometry, but the abundance of the proteins was relatively low. The instrumentation available at the time, namely the QSTAR-XL hybrid LC-MS/MS system, could not cope effectively with low abundance proteins present in such a complex mixture of peptides. As a result, protein identification was possible and the assigned total proteins scores, ranging from 20 to 200, were significant, albeit relatively low. Hence,

given this uncertainty regarding the validation of identified protein candidates, we decided to postpone publication of the results. However, can be seen herein a relatively high number (for such an endeavor) of samples were run and analyzed, adding weight nevertheless to the results. Consequently, it was decided to pursue this study when simultaneous availability of highly sensitive instrumentation, both for peptide separation and mass analysis, as well as fresh biological samples from CIAQ would not be a limitation. In the meantime, studies in the recent literature confirmed the presence of some of even the most unexpected candidates at the surface of mammalian spermatozoa, suggesting that the approach we took in 2003 is valid. This includes, for example, the presence of proteasone subunits at the surface of the spermatozoon. Hence, this reinforces our desires to further push this area of investigation and we have just undertaken the task of repeating this work using the modern proteomic facilities we have employed for the study reported in chapter 4. Once the total number of peptides and total protein scores will be judged sufficiently high to confidently discriminate between real and false-positive identifications, we will push further our analysis of the most promising candidates. Also, analysis of the identified proteins will include a search for PC4 cleavage motifs such as $KXKXXR^{\downarrow}$ or $KXXR^{\downarrow}$[36]. Early validation of the proposed substrates will be performed by monitoring their maturation (or absence of) in the mice PC4 knock-out model using Western blots.

3.9. FIGURES AND LEGENDS

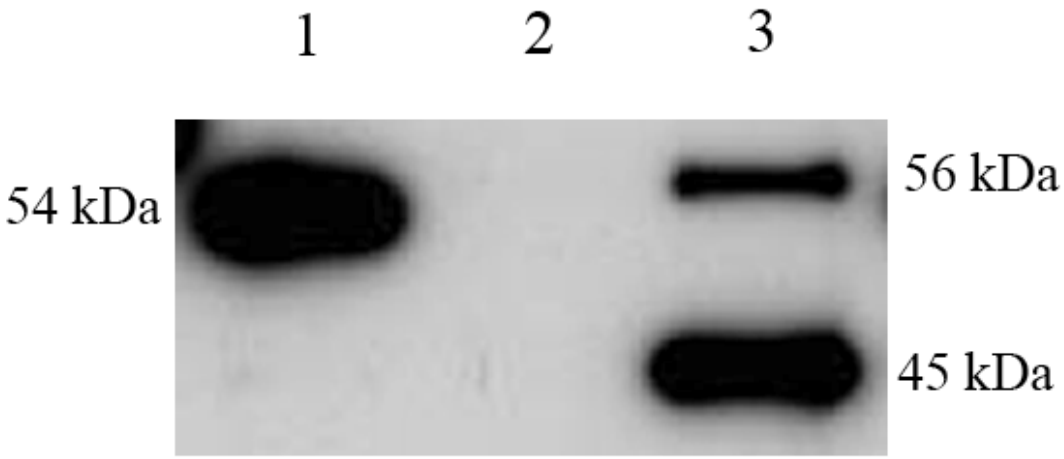


Figure 3.1: Representative Western blot illustrating the various PC4-like immunoreactive forms. Lane 1, supernatant of spermatozoa incubated in PBS for 15 minutes at 37°C; lane 2, consecutive pestle extract of spermatozoa in PBS; lane 3, consecutive pestle extract of spermatozoa in PBS containing 1% (v/v) NP-40 detergent. Proteins were revealed using a primary rabbit anti-PC4 antibody, a secondary anti-rabbit antibody conjugated to HRP and enhanced chemiluminescence. At least three distinct, sequentially extractable forms can be observed.

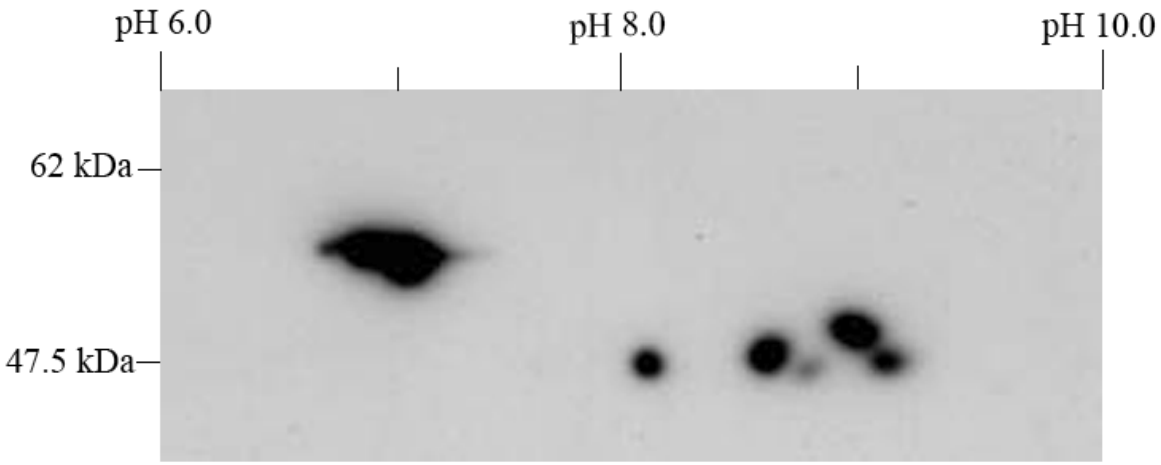


Figure 3.2: Western blot of PC4 immunoreactive proteins species separated by 2D-gel electrophoresis. Protein species were separated according to their pI (1st dimension, horizontal) and their molecular weight (2nd dimension, vertical) as described in Materials and methods. The observaiton of several spots indicates the presence of post-translational modification and/or multiple forms of the enzyme.

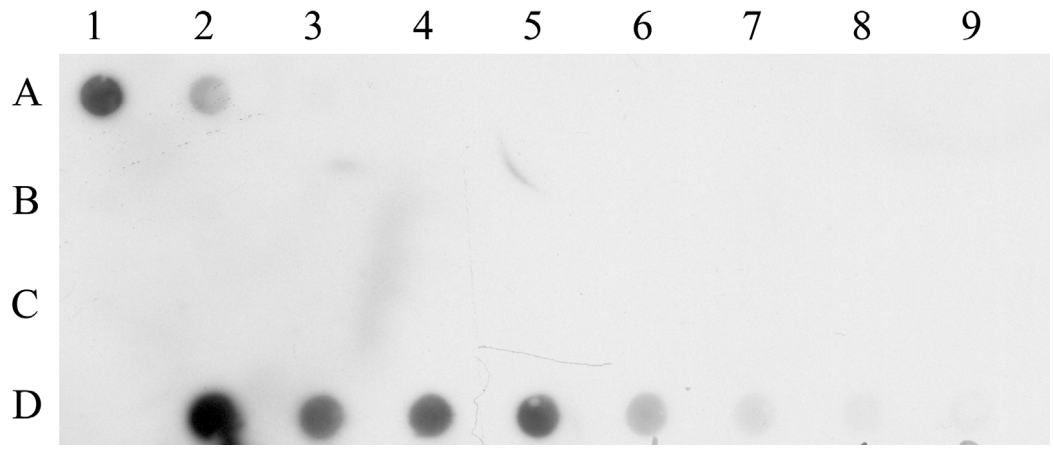


Figure 3.3: Dot blot illustrating the affinity purification of spermatozoa cell surface peptides. Row A, post-binding column wash 1, fractions 1-9; Row B, post-binding column wash 2, fractions 1-9; Row C, post-binding column wash 3, fractions 1-9; Row D, column eluate fractions 1-9. The presence of biotinylated peptides was revealed by neutravidin conjugated to HRP and enhanced chemiluminescence, as described.

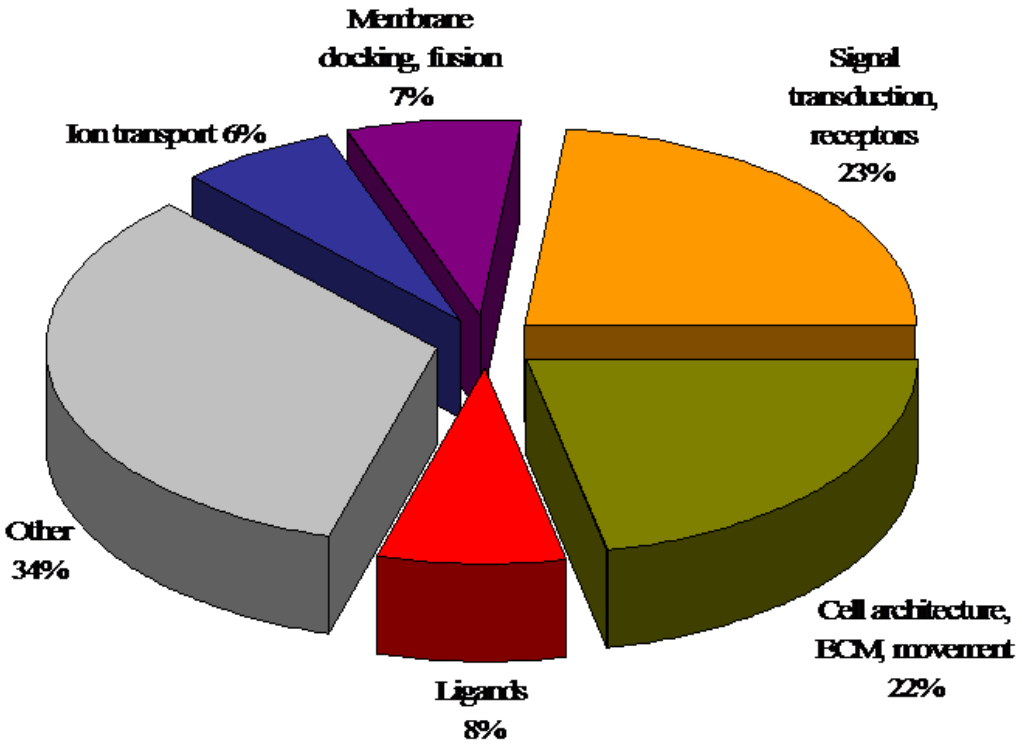


Figure 3.4: Distribution of the spermatozoa cell surface proteins identified by LC-MS/MS. The observed or postulated functions of the proteins confirm their transient or permanent association with the cellular membrane.

3.10 TABLES AND LEGENDS

Protein name	Function/Discovery
A-Kinase anchoring proteins	Cell structure and adhesion
ASPM (anti-sperm plasma membrane)	Sperm surface antigen
Collagen, type 7	Cell structure and adhesion
EP4 (low Mwt zona pellucida binding protein)	Zona pellucida binding
Glutathione peroxidase 5	Protection from lipid oxidation
Glutathione S-transferase	Protection from lipid oxidation
Hexokinase 1	Energy metabolism
Major Fibrous sheath protein precursor	Cell structure and adhesion
Mammalian 20S proteasome subunit	Zona pellucida binding, acrosome reaction
Mannose-6-phosphate/Insulin-like growth factor II receptor	Signaling in sperm development
PLC-β1 (phospholipase beta-1)	Acrosome reaction
Preprorelaxin precursor	Metabolism and acrosome reaction
RING-finger binding protein	Acrosome reaction
Ryanodine receptor	Ca^{2+} release in egg penetration
Sperm protein 13	Cell structure and adhesion

Table 3.1: **List of representative spermatozoa surface proteins identified by LC-MS/MS.** A few of the proteins identified are listed with their known or putative role in spermatozoa biology, or, if unavailable, the context in which they were initially described.

Supplementary Table 3.1

Protein Name	GI Number	Membrane	Model
Sample 1			
cardiac titin	gi|33413750	YES	NO
phosphatidylinositol-4-phosphate-5-kinase	gi|2072014	YES	NO
epithelial keratin 1	gi|50979272	YES	NO
EP4=low Mr zona pellucida binding protein	gi|254392	YES	YES
ryanodine receptor	gi|227245	YES	YES
Inositol 1,4,5-trisphosphate receptor type 1	gi|17367101	YES	YES
inositol 1,4,5-trisphosphate receptor type 3	gi|17432548	YES	NO
breast cancer susceptibility protein 2	gi|30060346	NO	NO
guanine nucleotide-binding protein Gs alpha1	gi|739576	YES	YES
sex determining region Y	gi|30016723	NO	YES
unconventional myosin	gi|47522864	YES	YES
Janus kinase 1	gi|47523036	YES	NO
ryanodine receptor	gi|11182064	YES	YES
ryanodine receptor	gi|164646	YES	YES
ASPM	gi|41056698	YES	YES
alpha 2B adrenergic receptor	gi|18643974	YES	NO
transcription factor NFATmac	gi|47523410	NO	NO
transferrin	gi|227094	YES	YES
mitosin	gi|12002984	NO	NO
desmoglein 3	gi|50950209	YES	NO
tCstF-64	gi|24416589	NO	YES
MHC class I allele ShB	gi|424077	YES	NO
smooth muscle caldesmon	gi|845652	NO	NO
microcephalin	gi|46399208	NO	NO
organic anion transporter 3	gi|42538740	YES	YES
Phosphorylated Pig Heart, Gtp-Specific Succinyl-Coa Synthetase	gi|9955021	NO	NO
cardiac titin	gi|33413748	YES	NO
Clp-like ATP-dependent protease binding subunit	gi|871784	NO	NO
BRCA1	gi|18766288	NO	YES
Glutathione S-transferase	gi|2495107	YES	YES
signal recognition particle,72 kDa subunit	gi|50979066	YES	NO
DNA (cytosine-5)-methyltransferase 1	gi|29536011	NO	YES
BRCA1	gi|32892158	NO	YES
Sample 2			
ASPM	gi|41056696	YES	YES
BRCA1	gi|13195179	NO	YES
Cleavage and polyadenylation specificity factor, 73 kDa subunit	gi|18202362	NO	NO
Pyruvate Kinase	gi|999572	NO	YES
cytosine-5-methyltransferase	gi|22023943	NO	YES

Hexokinase, type I	gi|123891	YES	YES
collagen alpha1(II)	gi|222999	YES	YES
inversin	gi|50979224	NO	NO
transforming growth factor beta receptor type II	gi|348620	YES	YES
ATP-binding cassette, sub-family F (GCN20), member 1	gi|41529180	NO	NO
corneal epithelium BCP54	gi|162904	NO	NO
NADH dehydrogenase	gi|163412	NO	NO
vinculin	gi|47522618	YES	YES
RING-finger binding protein	gi|30315951	YES	YES
myosin heavy chain 2a	gi|5360746	YES	NO
Junctophilin 1	gi|27805491	YES	YES
Myosin heavy chain, nonmuscle type B	gi|13431706	YES	NO
pulmonary surfactant-associated protein C proSP-C	gi|13507047	YES	NO
60S Ribosomal protein L34	gi|2833354	NO	NO
Chromaffin granule-associated membrane glycoprotein IIA	gi|1170800	YES	NO
type VII collagen	gi|33149359	YES	YES
fibrinogen A-alpha chain	gi|3789960	YES	NO
procollagen alpha1(IV)	gi|1915982	YES	NO
putative zinc finger protein	gi|6625537	NO	NO
Vitamin D-binding protein precursor (DBP) (GC-globulin) (VDB)	gi|1722804	YES	YES
versican precursor	gi|7513547	YES	NO
Cholinesterase precursor	gi|38502852	YES	YES
ribosome receptor	gi|50978924	YES	NO
adult-specific brush border esterase/phospholipase	gi|539770	YES	YES
Interphotoreceptor retinoid-binding protein precursor	gi|124891	YES	NO
BRCA1	gi|25070848	NO	YES
Integrin beta-2 precursor/ CD18	gi|417199	YES	YES
lysosomal trafficking regulator	gi|4757355	YES	NO
12-lipoxygenase	gi|16212141	YES	NO
capacitative calcium entry channel 2	gi|3212137	YES	YES
serum albumin precursor	gi|50953792	NO	NO
T-cell receptor delta 3	gi|11071821	YES	NO
complement component C5	gi|48675959	YES	YES
mannose-6-phosphate/insulin-like growth factor II receptor	gi|14647147	YES	YES
antibacterial protein precursor	gi|468912	YES	YES
met proto-oncogene precursor	gi|38322673	YES	YES
mannose 6-phosphate/insulin-like growth factor 2 receptor	gi|7715870	YES	YES
Protein disulfide-isomerase A3 precursor er-60	gi|729433	YES	NO
DNA-dependent protein kinase catalytic subunit	gi|17646639	NO	NO
insulin-like growth factor-binding protein 1 precursor	gi|108549	YES	NO
Myosin light chain kinase 2, skeletal/cardiac muscle	gi|125493	YES	YES
alpha 3 actinin	gi|47155482	NO	NO
Glucose transporter type 4, insulin-responsive	gi|2500935	YES	NO
Plakophilin	gi|599690	YES	NO

Sample 3			
Versican core protein precursor	gi|19861780	YES	NO
Voltage-dependent N-type calcium channel alpha-1B subunit	gi|5921693	YES	NO
interphotoreceptor retinoid binding protein	gi|2570813	YES	NO
heavy neurofilament protein; intermediate filament protein	gi|50979202	NO	YES
interphotoreceptor retinoid binding protein	gi|22796553	YES	NO
Common salivary protein form a BSP30a	gi|34395843	NO	NO
breast and ovarian cancer susceptibility protein	gi|29640734	NO	YES
interphotoreceptor retinoid binding protein	gi|34732754	YES	NO
phosphoprotein phosphatase 2A-alpha 65K regulatory chain	gi|108375	YES	YES
Translation initiation factor IF-2, mitochondrial precursor	gi|1170498	NO	NO
von Willebrand factor (homolog?)	gi|17980666	NO/YES	NO/YES
heat shock protein 90	gi|1072479	NO	YES
alpha-1D adrenergic receptor	gi|6114881	YES	NO
chromogranin A precursor	gi|89468	NO	NO
titin	gi|1145880	YES	NO
tropomyosin beta chain, platelet	gi|71604	YES	YES
aromatase	gi|1762231	NO	YES
periplakin	gi|20149233	YES	NO
protein phosphatase 1, regulatory subunit 10	gi|41529181	NO	YES
BRCA1	gi|13195215	NO	YES
Nuclear autoantigenic sperm protein	gi|127843	NO	YES
calcium channel BI-1	gi|1523	YES	NO
Sample 4			
sperm protein 13	gi|40644925	YES	YES
Na+ Ca2+ K+-exchanging protein	gi|108825	YES	NO
brain calcium channel BII-1	gi|1473	YES	NO
Isonicotinimidylated Liver Alcohol Dehydrogenase	gi|231239	NO	NO
tuftelin	gi|27530964	YES	NO
BRCA1	gi|18766276	NO	YES
myosin heavy chain, smooth muscle, long splice form	gi|109322	YES	YES
sarcalumenin precursor, fast twitch skeletal muscle	gi|109375	YES	NO
interleukin 12 receptor beta 2 chain	gi|18139969	YES	NO
BRCA1	gi|13195205	NO	YES
smoothelin-B	gi|45154853	NO	NO
dentin matrix protein 1	gi|26000384	YES	NO
PLC-BETA-1	gi|130221	YES	YES
Rho-associated protein kinase 2	gi|47606001	YES	NO
Rho-associated coiled-coiled containing protein kinase p160ROCK	gi|21306504	YES	NO
DNA cytosine-5 methyltransferase 1	gi|31074161	NO	NO
cytochrome P450 2D	gi|50979325	NO	NO
aggrecan	gi|1730260	YES	NO

adhesion molecule VCAM-1	gi|28894262	YES	NO
growth hormone receptor	gi|21645703	YES	YES
Ryanodine receptor 1	gi|134134	YES	YES
BRCA1	gi|18766284	NO	YES
myomesin	gi|2136688	NO	NO
Ly49	gi|30267428	YES	NO
dentin matrix protein 1	gi|26000368	YES	NO
immunoglobulin kappa light chain	gi|30269858	YES	NO
Collagen alpha 1(III) chain	gi|115290	YES	YES
TPA: RTN2-B	gi|38327596	YES	NO
cardiac ankyrin repeat protein	gi|47522642	YES	NO/YES
cortactin-binding protein 2	gi|38322680	YES	NO
porcine orexin receptor type 1	gi|20520985	YES	NO
dentin matrix protein 1	gi|26000372	YES	NO
calmodulin-independent adenylate cyclase	gi|61	YES	NO
DNA-(apurinic or apyrimidinic site) lyase (AP endonuclease 1) (APEX nuclease) (APEN)	gi|113983	NO	NO
troponin T fast skeletal muscle type	gi|21038992	NO	NO
Sodium/calcium exchanger 1 precursor (Na(+)/Ca(2+)-exchange protein 1)	gi|127793	YES	YES
Sample 5			
ORF	gi|164251	NO	NO
epithelial keratin 2e	gi|50979264	YES	NO
p19ARF	gi|7637746	NO	NO
putative collagen type XI alpha 1	gi|27657427	YES	NO
skeletal muscle sodium channel alpha-subunit	gi|829615	YES	NO
interleukin-1 beta precursor	gi|69700	YES	YES
follicle-stimulating hormone receptor	gi|1658014	YES	NO
Brca2	gi|20502478	NO	NO
titin	gi|109388	YES	NO
Hyaluronan synthase type 2	gi|3927898	YES	NO
cortactin-binding protein 2	gi|38322718	YES	NO
uveal autoantigen	gi|12240158	YES	NO
lymphocyte function-associated antigen 1	gi|24637288	YES	NO
Nitric-oxide synthase, brain	gi|8473494	YES	YES
protein 4.1R	gi|50979218	YES	NO
Calpain inhibitor (Calpastatin)	gi|124110	NO	YES
BRCA1	gi|13195173	NO	YES
BRCA1	gi|18874718	NO	YES
Neurofilament triplet M protein	gi|1709261	NO	YES
zinc finger protein 470	gi|45439038	NO	NO
putative 40-2-3 protein	gi|47564034	UNK	UNK
thioredoxin reductase 2	gi|27807131	NO	NO
ATPase inhibitor, mitochondrial precursor	gi|2833356	NO	NO
BRCA1	gi|32892174	NO	YES
xeroderma pigmentosum group C protein	gi|3098416	NO	NO

xenobiotic/medium-chain fatty acid:CoA ligase form XL-III	gi|5070357	NO	NO
pancreatic colipase	gi|50979104	NO	NO
parotid secretory protein precursor	gi|33358232	NO	NO
Collagen alpha 1(VIII) chain precursor	gi|115317	YES	YES
reductase,NADPH cytochrome P450	gi|224382	YES	NO
Sample 6			
major fibrous sheath protein precursor	gi|4588120	YES	YES
dystrophin	gi|3982751	YES	YES
acylneuraminate lyase	gi|47522956	YES	YES
antigen CD18	gi|45268485	YES	YES
chmadrin (short type)	gi|4996101	NO	NO
Coatomer alpha subunit (Alpha-coat protein) (Alpha-COP)	gi|2494888	YES	YES
immunoglobulin kappa light chain variable region	gi|6502690	YES	NO
recombination activating protein 1	gi|30314366	NO	NO
BRCA1	gi|23928361	NO	YES
insulin receptor precursor	gi|37813344	YES	YES
Probable phospholipase DDHD1	gi|37999475	YES	YES
BRCA1	gi|18766286	NO	YES
glutathione peroxidase 5	gi|6650995	YES	YES
IkB kinase-gamma	gi|15986413	NO	NO
interphotoreceptor matrix proteoglycan 1	gi|27805947	YES	NO
Endoplasmin precursor	gi|984249	YES	YES
Cyclic-nucleotide-gated cation channel 4	gi|2493749	YES	NO
blood coagulation factor IX	gi|45758733	NO	NO
myosin heavy chain 2x	gi|21907900	YES	NO
BRCA1	gi|13991595	NO	YES
kallikrein	gi|47522962	YES	YES
huntingtin interacting protein 1-related	gi|32699406	YES	NO
cardiac triadin isoform 2	gi|1144323	YES	NO
apolipoprotein B	gi|32895321	YES	YES
Glutamate decarboxylase, 67 kDa isoform (GAD-67)	gi|416884	NO	YES
prostaglandin F2-alpha receptor	gi|15421164	YES	YES/NO
major histocompatibility complex class I protein	gi|6049084	YES	NO
interleukin-3	gi|45359670	YES	YES
histone H2A.F/Z variant	gi|3108211	NO	NO
block of proliferation 1	gi|33411760	UNK	UNK
sialyltransferase ST3Gal-III	gi|50872129	YES	NO
destrin	gi|5802966	YES	YES
Sample 7			
Versican core protein precursor (Large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (PG-M)	gi|19861780	YES	YES
retinitis pigmentosa GTPase regulator	gi|9837383	UNK	NO
immunoglobulin heavy chain VHDJ region	gi|38092440	YES	NO

Senescence marker protein-30 (SMP-30) (Regucalcin) (RC)	gi|13633941	NO	NO
1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (PLC)	gi|108854	YES	YES
Pyruvate kinase, isozyme M1	gi|125598	NO	YES
Clathrin heavy chain	gi|1705915	YES	YES
A-kinase anchoring protein AKAP120	gi|3599584	YES	YES
Phosphorylase B kinase alpha regulatory chain	gi|125531	NO	NO
Neurofilament triplet M protein (160 kDa neurofilament protein)	gi|13629976	NO	YES
BRCA1	gi|18874724	NO	YES
pyruvate kinase; ATP:pyruvate 2-o-phosphotransferase	gi|2623945	NO	YES
Chloride intracellular channel 6 (Parchorin)	gi|24211556	YES	NO
bovine submaxillary mucin 1	gi|3057087	YES	NO
Chain D, Succinyl-Coa:3-Ketoacid Coa Transferase	gi|28373592	NO	YES
POU domain, class 5, transcription factor 1	gi|28201856	NO	NO
Chain B, Structure Of The C Domain Of Synapsin Ia	gi|2981708	YES	NO
NADPH-cytochrome P-450 oxidoreductase	gi|499862		
Sample 8			
ryanodine receptor type 3	gi|1526615	YES	YES
myosin heavy chain 2a	gi|21907898	YES	NO
protein tyrosine kinase fer	gi|50978868	NO	YES
PREPRORELAXIN PRECURSOR	gi|20454199	YES	YES
transferrin	gi|7513541	YES	YES
ADP-ribosyl cyclase (Cyclic ADP-ribose hydrolase) (CD38 homolog)	gi|21263460	YES	NO
Tenascin-X	gi|2462979	YES	YES
Bucentaur	gi|2114146	NO	YES
Pigment epithelium-derived factor, PEDF	gi|1176450	YES	NO
alpha adrenergic receptor 2B	gi|32140145	YES	NO
apolipoprotein B 100	gi|33150951	YES	YES
guanylate cyclase 70 kDa subunit	gi|408	NO	NO
KIAA0373	gi|6007639	UNK	UNK
Gag	gi|6651075	NO	NO
ferret 5-HT3A receptor	gi|8247752	YES	NO
factor VIII	gi|2645493	YES	NO
dentin matrix protein 1	gi|26000374	YES	NO
interleukin-12 p40 subunit	gi|984511	YES	YES
Myc proto-oncogene protein	gi|2498005	NO	YES
inter-alpha-inhibitor heavy-chain 1	gi|1915956	YES	NO
synaptic nuclear envelope protein	gi|32141360	YES	NO
calcyclin-associated protein, CAP50=Ca2+/phospholipid-binding protein L-16 fragment	gi|249043	YES	NO
vitamin K-dependent gamma-glutamyl carboxylase	gi|8927574	YES	YES
Sialin (Solute carrier family 17 member 5) (Sodium/sialic acid cotransporter) (Membrane glycoprotein SP55)	gi|48428686	YES	NO

CD2 antigen	gi|40363613	YES	NO
Y-linked zinc finger protein	gi|16416473	NO	YES
NFBD1	gi|41529174	NO	NO
neurocalcin gamma 1	gi|1072436	YES	NO
chromogranin A	gi|198	NO	NO
Sodium/myo-inositol cotransporter	gi|1709224	YES	NO
Tenomodulin	gi|23200577	YES	YES
Structural maintenance of chromosome 1-like 1 protein (SMC-protein)	gi|29336595	NO	NO
ataxia-telangiectasia mutated protein	gi|46810446	NO	YES
Brevican core protein precursor	gi|2507585	YES	NO
myosin heavy chain 2b	gi|5360748	YES	NO
steroid 11beta-monooxygenase	gi|89771	NO	NO
transient receptor potential V1	gi|50236426	YES	NO
Mammalian 20s Proteasome	gi|21465662	YES	YES
phosphoprotein phosphatase (EC 3.1.3.16) 1 glycogen-binding regulatory chain	gi|109363	NO	YES
alpha-fetoprotein	gi|47523700	YES	YES
Neurofilament triplet L protein	gi|2506775	NO	YES
5-oxo-L-prolinase	gi|46020038	NO	NO
MHC II DR-beta 1 chain	gi|2326247	YES	NO
TPA: RTN4-Aw	gi|38488991	YES	NO
Fibrillin 1 precursor (MP340)	gi|1706768	YES	NO
CD40 ligand	gi|1083014	YES	NO

Supplementary Table 3.1: Spermatozoa surface proteins identified by LC-MS/MS. This table lists all proteins identified from all 8 samples analyzed, along with their gi identification number. It also indicates if the protein was previously reported as being cell surface associated, and if it was reported as being present in this model, the mammalian spermatozoa.

3.11 BIBLIOGRAPHY

1. Seidah, N.G. et al. Testicular expression of PC4 in the rat: molecular diversity of a novel germ cell-specific Kex2/subtilisin-like proprotein convertase. *Mol. Endocrinol.* **6**, 1559-1570 (1992).

2. Mbikay, M. et al. Impaired fertility in mice deficient for the testicular germ-cell protease PC4. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6842-6846 (1997).

3. Pesch, S. & Bergmann, M. Structure of mammalian spermatozoa in respect to viability, fertility and cryopreservation. *Micron* **37**, 597-612 (2006).

4. Visconti, P.E. et al. The molecular basis of sperm capacitation. *J. Androl.* **19**, 242-248 (1998).

5. Visconti, P.E. & Kopf, G.S. Regulation of protein phosphorylation during sperm capacitation. *Biol. Reprod.* **59**, 1-6 (1998).

6. Rufo, G.A., Jr., Schoff, P.K. & Lardy, H.A. Regulation of calcium content in bovine spermatozoa. *J. Biol. Chem.* **259**, 2547-2552 (1984).

7. Rochwerger, L. & Cuasnicu, P.S. Redistribution of a rat sperm epididymal glycoprotein after in vitro and in vivo capacitation. *Mol. Reprod. Dev.* **31**, 34-41 (1992).

8. Cowan, A.E., Koppel, D.E., Vargas, L.A. & Hunnicutt, G.R. Guinea pig fertilin exhibits restricted lateral mobility in epididymal sperm and becomes freely diffusing during capacitation. *Dev. Biol.* **236**, 502-509 (2001).

9. Harkema, W., Harrison, R.A., Miller, N.G., Topper, E.K. & Woelders, H. Enhanced binding of zona pellucida proteins to the acrosomal region of the intact boar spermatozoa in response to fertilizing conditions: a flow cytometric study. *Biol. Reprod.* **58**, 421-430 (1998).

10. Nixon, B. et al. The identification of mouse sperm-surface-associated proteins and characterization of their ability to act as decapacitation factors. *Biol. Reprod.* **74**, 275-287 (2006).

11. Sirivaidyapong, S., Bevers, M.M. & Colenbrander, B. Acrosome reaction in dog sperm is induced by a membrane-localized progesterone receptor. *J. Androl.* **20**, 537-544 (1999).

12. Abou-Haila, A. & Tulsiani, D.R. Mammalian sperm acrosome: formation, contents, and function. *Arch. Biochem. Biophys.* **379**, 173-182 (2000).

13. Snell, W.J. & White, J.M. The molecules of mammalian fertilization. *Cell* **85**, 629-637 (1996).

14. Flesch, F.M. & Gadella, B.M. Dynamics of the mammalian sperm plasma membrane in the process of fertilization. *Biochim. Biophys. Acta* **1469**, 197-235 (2000).

15. Geng, J.G. et al. Expression of a P-selectin ligand in zona pellucida of porcine oocytes and P-selectin on acrosomal membrane of porcine sperm cells. Potential implications for their involvement in sperm-egg interactions. *J. Cell Biol.* **137**, 743-754 (1997).

16. Mori, E., Baba, T., Iwamatsu, A. & Mori, T. Purification and characterization of a 38-kDa protein, sp38, with zona pellucida-binding property from porcine epididymal sperm. *Biochem. Biophys. Res. Commun.* **196**, 196-202 (1993).

17. Richardson, R.T. & O'Rand, M.G. Site-directed mutagenesis of rabbit proacrosin. Identification of residues involved in zona pellucida binding. *J. Biol. Chem.* **271**, 24069-24074 (1996).

18. Hunnicutt, G.R., Primakoff, P. & Myles, D.G. Sperm surface protein PH-20 is bifunctional: one activity is a hyaluronidase and a second, distinct activity is required in secondary sperm-zona binding. *Biol. Reprod.* **55**, 80-86 (1996).

19. Forsbach, A. & Heinlein, U.A. Intratesticular distribution of cyritestin, a protein involved in gamete interaction. *J. Exp. Biol.* **201**, 861-867 (1998).

20. Primakoff, P., Hyatt, H. & Tredick-Kline, J. Identification and purification of a sperm surface protein with a potential role in sperm-egg membrane fusion. *J. Cell Biol.* **104**, 141-149 (1987).

21. Nakagawa, T. et al. Identification and functional expression of a new member of the mammalian Kex2-like processing endoprotease family: its striking structural similarity to PACE4. *J. Biochem. (Tokyo).* **113**, 132-135 (1993).

22. Seidah, N.G. & Prat, A. Precursor convertases in the secretory pathway, cytosol and extracellular milieu. *Essays Biochem.* **38**, 79-94 (2002).

23. Gyamera-Acheampong, C. et al. Sperm from mice genetically deficient for the PCSK4 proteinase exhibit accelerated capacitation, precocious acrosome reaction, reduced binding to egg zona pellucida, and impaired fertilizing ability. *Biol. Reprod.* **74**, 666-673 (2006).

24. Li, M., Mbikay, M. & Arimura, A. Pituitary adenylate cyclase-activating polypeptide precursor is processed solely by prohormone convertase 4 in the gonads. *Endocrinology* **141**, 3723-3730 (2000).

25. Li, M., Funahashi, H., Mbikay, M., Shioda, S. & Arimura, A. Pituitary adenylate cyclase activating polypeptide-mediated intracrine signaling in the testicular germ cells. *Endocrine* **23**, 59-75 (2004).

26. Martinez-Heredia, J., Estanyol, J.M., Ballesca, J.L. & Oliva, R. Proteomic identification of human sperm proteins. *Proteomics* **6**, 4356-4369 (2006).

27. Li, L.W. et al. Establishment of a high-resolution 2-D reference map of human spermatozoal proteins from 12 fertile sperm-bank donors. *Asian J. Androl.* **9**, 321-329 (2007).

28. Stein, K.K., Go, J.C., Lane, W.S., Primakoff, P. & Myles, D.G. Proteomic analysis of sperm regions that mediate sperm-egg interactions. *Proteomics* **6**, 3533-3543 (2006).

29. Cleveland, D.W., Fischer, S.G., Kirschner, M.W. & Laemmli, U.K. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* **252**, 1102-1106 (1977).

30. Gauthier, D.J., Gibbs, B.F., Rabah, N. & Lazure, C. Utilization of a new biotinylation reagent in the development of a nondiscriminatory investigative approach for the study of cell surface proteins. *Proteomics* **4**, 3783-3790 (2004).

31. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

32. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

33. Mbikay, M. et al. Structure of the gene for the testis-specific proprotein convertase 4 and of its alternate messenger RNA isoforms. *Genomics* **20**, 231-237 (1994).

34. Pasten, C., Morales, P. & Kong, M. Role of the sperm proteasome during fertilization and gamete interaction in the mouse. *Mol. Reprod. Dev.* **71**, 209-219 (2005).

35. Chiarella, P., Puglisi, R., Sorrentino, V., Boitani, C. & Stefanini, M. Ryanodine receptors are expressed and functionally active in mouse spermatogenic cells and their inhibition interferes with spermatogonial differentiation. *J. Cell Sci.* **117**, 4127-4134 (2004).

36. Basak, S., Chretien, M., Mbikay, M. & Basak, A. In vitro elucidation of substrate specificity and bioassay of proprotein convertase 4 using intramolecularly quenched fluorogenic peptides. *Biochem. J.* **380**, 505-514 (2004).

Chapter 4

Organellar proteomics in a day: Flow cytometry-assisted purification and proteomic analysis of the corticotropes dense-core secretory granules

Daniel J GAUTHIER[1], Jacqueline A. SOBOTA[2], Francesco FERRARO[2], Richard E MAINS[2] & Claude LAZURE[1]

[1]Neuropeptides Structure & Metabolism Research Unit, Institut de recherches cliniques de Montréal (affiliated to University of Montréal), 110 Pine avenue West, Montréal, Québec H2W 1R7, Canada.

[2]Department of Neuroscience, University of Connecticut Health Center, Farmington, CT 06030-3401.

Manuscript submitted to *Proteomics*.

4.1 PREFACE

Interestingly, despite significant technological advances in mass spectrometry-based protein identification, one aspect of subcellular proteomics, namely the purification of organelles, still relies mostly on classical methods. Effectively, density gradient centrifugation of whole cell lysates is still in most cases the method of choice to enrich a sample for a specific subcellular structure. Among the main disadvantages of this procedure are the variable degree of enrichment, the relatively large amounts of biological material required, the laborious manipulations and last, but not least, the duration of the complete procedure. On average, in subcellular proteomic studies employing density gradient centrifugation, between 48 and 72 hours are elapsed between cell or tissue lysis and MS data acquisition. Considering the fact that organelles are complex and dynamic structures, maintaining them at 4°C for such a prolonged duration could lead to an alteration in their protein composition. On this topic, Yates III *et al.* recently commented that "(organelle) separation methods that give a higher resolution would facilitate the analysis of all cellular components using proteomics technologies, so the development of new methods is a challenge for the field**."**[1] In order to simplify, improve and accelerate the purification of organelles, we have decided to develop a completely different approach to organelle purification based on fluorescence-assisted sorting. This concept, although first proposed in 1985[2], has remained relatively unexplored since. In this chapter, we describe the successful fluorescence-assisted purification of corticotropes secretory granules and their proteomic analysis within a single day of work. The rapidity, efficiency and adaptability of this gel-free procedure contrast sharply with the protocols currently used in the field of proteomics. The quality

of the results presented here results from months of development and optimization of FACS and proteomics techniques, and the very recent publication of proteomes from similar organelles suggest that it is superior to classical procedures in several aspects including the degree of purification, the score of identified proteins and the identification of proteins transiently associated to the granules. In a way, we propose our response to the above-mentioned challenge in the field of subcellular proteomics.

4.2 SUMMARY

The field of organellar proteomics has emerged as an attempt to minimize the complexity of the proteomics data obtained from whole cell and tissue extracts while maximizing the resolution on the protein composition of a single subcellular compartment. Standard methods involve lengthy density-based gradient and/or immunoaffinity purification steps followed by extraction, one-dimensional or two-dimensional gel electrophoresis, gel staining, in-gel tryptic digestion and protein identification by mass spectrometry. In this paper, we present a simple approach to purify endocrine secretory granules and obtain a proteome of over 150 proteins in a single day. The gel-free procedure involves fluorescence-assisted sorting of the secretory granules followed by gentle extraction in a buffer compatible with tryptic digestion and mass-spectrometry. This procedure requires no modification to any instrumentation and can be readily adapted to the study of other organelles.

4.3 INTRODUCTION

As a whole, the field of proteomics aims to identify all proteins present in a tissue or cell type in a given set of conditions at a specific time. However, this endeavor has proven extremely complex due to the thousands of proteins being potentially expressed at a dynamic level varying from a few copies to millions of copies within a single cell.

In an effort to circumvent the ensuing data complexity, a subfield of proteomics has emerged and proposes analysis of the proteome of individual subcellular compartments. In the last few years, numerous studies have been published and led to the identification of anywhere between one hundred to a few thousand proteins specific to many organelles[1, 3], including the cell nucleus in terms of its content in the nuclear envelope[4] and the nucleolus[5], endoplasmic reticulum[6], the Golgi apparatus[7], the secretory lysosomes[8], the phagosome[9], clathrin-coated vesicles[10] and the mitochondrion[11]. The resulting information has been of the utmost importance in unraveling the specific overlapping and non-overlapping protein composition of cellular compartments as well as their specific roles in biological processes.

In nearly all the subcellular proteomics studies including those above listed, the general methodology relies to a great extent on subcellular fractionation allowing recovery of the organelle through careful selection of physical properties. Hence, following cell or tissue lysis, it is usual to submit the lysate to one or many rounds of differential centrifugation. Velocity and equilibrium centrifugation have established their capability in the past fifty years to separate cellular components differing by density. Once the centrifugation steps are complete, the organellar content is extracted and the proteins are further separated by 1D- or 2D-SDS-PAGE. Upon gel staining, the bands or spots are excised; the proteins are proteolytically digested and are ultimately identified by

analyzing their peptides by mass spectrometry. However, while adequate for tissular cell extraction, this entire lengthy procedure (48-72 hrs) is rendered less efficient when applied to cell cultures[12].

In addition to the difficulty and the duration of the procedure, we aimed at simplifying it for two other main reasons. Firstly, it is still unclear how representative the proteome of an organelle is when taking into perspective the time at which cells were extracted and the time at which MS data is made available following days of preparation, even in the presence of protease inhibitors. Secondly, even after optimization, most steps involved are not very efficient, thereby necessitating large amounts of starting biological material. For example, SDS-PAGE leads unavoidably to staining artifacts, loss of sample and poor separation of very high (over 100 kDa) and very low (under 15 kDa) molecular weight proteins.

Fluorescence assisted cell sorting (FACS) is routinely employed to perform selective enrichment of cell populations bearing specific markers based on fluorescence[13, 14]. In 1985, Murphy's group first described the concept of single organelle fluorescence analysis (SOFA) and subsequently suggested its application to sort single organelles[2]. The term fluorescence-assisted organelle sorting (FAOS) has been introduced to describe the use of flow cytometry to sort subcellular structures and components. However, this represents a challenging task as the organelles to be sorted are routinely 10 to 100 times smaller than the original cells. Furthermore, the resistance of the target organelle to the shear and tear fluid forces in the cell sorter is limited. Nonetheless, publications have described sorting of organelles such as endosomes[15], mitochondria[16] and phagosomes[17] using various fluorescent probes. However, due to technical constraints, most of these were analytical in nature and none went as far as conducting a proteomic analysis of the

sorted organelle although demonstration of significant level of enrichment was demonstrated[18].

The dense-core secretory granule is the ultimate compartment in the regulated pathway of secretion. Originally perceived as a simple storage compartment, it is now known to be actively involved in cargo selection, maturation, and secretory processes. However, much remains to be defined in terms of the mechanisms implicated and the proteins involved. Proteomics studies of the secretory pathway[6] as well as that of various secretory vesicles and/or granules have been published, some very recently. Using classical purification and analysis protocols, the proteomes of atrial secretory granules[19], pancreatic zymogen granules[20], chromaffin granules[21], insulin-containing granules[22] and synaptic vesicles[23, 24] were documented.

This work describes the application of a novel methodology to the study of the pituitary corticotropes dense-core secretory granules which does not use density gradient centrifugation and gel separation. It involves gentle cell lysis, brief centrifugation, improved FAOS and LC-MS. Using relatively small numbers of cultured cells, it led to identification of 158 proteins, including small molecular weight proteins and transiently granule-associated proteins, with less contamination arising from many subcellular compartments. More importantly, it can be accomplished within a day following proper cell culturing and using readily available facilities.

4.4 EXPERIMENTAL PROCEDURES

4.4.1 Cell culture and immunocytochemistry

4.4.1.1 Materials

Unless otherwise stated, all chemicals, solutions and solvents were of the highest possible grade purchased from Sigma-Aldrich. Primary antibodies (anti-calnexin, anti-giantin, anti-ACTH, anti-prohibitin, anti-ERp29, anti-PIST, anti-GFP were purchased from Abcam. Cell culture material was purchased from Invitrogen.

4.4.1.2 Cell culture and immunocytochemistry

The AtT-20 cells stably expressing the-PHM-mGFP fusion protein were generated as described[25]. Briefly, AtT-20 were transfected with an expression vector encoding for amino acids 1-407 of the peptidyl-glycine α-amidating monooxygenase enzyme fused in frame with monomeric green fluorescent protein. AtT-20 and AtT-20-PHM-mGFP cells were grown in DMEM-F12 containing 10% fetal calf serum, 10% NuSerum, penicillin/streptomycin/glutamine and 0.5 mg/mL G418 in T175 flasks at 37ºC in 10% CO_2 atmosphere. They were regularly passaged when confluence reached 90-95%. Fresh medium was added to the cells 16-20 hrs before they were collected. Cells were usually kept for no more than 20 generations in order to maintain phenotype and steady levels of fluorescence. WT and PHM-mGFP transfected cells were indistinguishable in appearance and growth properties. Colocalization of ACTH with PHM-mGFP in the secretory

granules was seen using a Zeiss LSM510 confocal microscope following procedures previously described[25].

4.4.2 Cell lysis and fluorescence-assisted organelle sorting (FAOS)

4.4.2.1 Cell lysis

All the following steps were performed at 37ºC. When cells reached 90-95% confluence, they were trypsinized and collected in warm fresh medium. Latrunculin B and nocodazole (Sigma-Aldrich) were added to a final concentration of 5 µM each and the cells were allowed to incubate for 20 min with occasional gentle agitation. The cells from a single T175 flask were pelleted, washed once in PBS and resuspended in 1 mL of sucrose solution (0.34 M sucrose, protease inhibitor cocktail (Roche), 5 µM latrunculin B, 5 µM nocodazole). From this point on, all steps were carried at 4ºC. Two cell lysis procedures were evaluated namely the use of a 2 mL Potter homogenizer (Kontes) and the use of a freeze-thaw procedure. However, submitting them to one freeze (5 min)-thaw (2 min) cycle consistently resulted in over 90% cell lysis as seen under light microscope. The cell extract was further incubated 5 min on ice, and then centrifuged on a bench-top centrifuge at 1 500 x g for 5 min to remove nuclei, unbroken cells and large debris. The 1 500 x g supernatant was collected and diluted with one third volume of cold PBS with EDTA (25 mM final). This final sample was immediately submitted to flow cytometry.

4.4.2.2 FAOS

Flow cytometry sorting of the secretory granules was performed using a MOFLO ultra high speed cell sorter (Dako) equipped with an Innova 90C ion laser (Coherent)

tuned at 488nm and 200mW. Threshold was adjusted on the side scatter (SSC) set in linear mode. Green florescence was detected in FL1 using standard filters for GFP detection (530/30). The threshold was adjusted to optimize granules detection, in which case the sorting rate was in the range of 3 000-5 000 events per second, even if over 25 000 events per second were detectable at maximum sensitivity. Since this equipment is usually employed to sort whole cells (6-25 μm), the neutral filter was removed to increase sensitivity in the lower size range and the sample pressure was set to 20 psi (1.38 bars). In order to define reference settings, 450 nm sky blue, 530 nm nile red, 840 nm sky blue and 840 nm nile red control beads (Spherotech) were used to obtain a size and fluorescence estimate of the sorted events. All sorted events were smaller than 500 nm, as fluorescent as fluorescent nile red 530 nm beads, and at least 10 times more fluorescent than non-fluorescent sky blue 450 nm beads. Non-transfected WT AtT-20 cells were also analyzed and results confirmed the absence of endogenous green fluorescence. Propidium iodide staining was routinely conducted to confirm the absence of DNA left in the samples after centrifugation and before sorting to exclude nuclei. Data acquisition and processing was performed with the Summit software, v4.3 (Dako). In all cases, the content of the cell lysate was sorted in ice-cold PBS containing 5 μM latrunculin B and the purified granules were centrifuged at 5 000 x g for 20 min at 4ºC using a swing-out SW41ti rotor in a L8-80 Ultracentrifuge (Beckman). Following removal of the supernatant, the pellet was washed with cold PBS and centrifuged again at 5 000 x g for 20 min at 4ºC.

4.4.3 Electron microscopy and Western Blot

4.4.3.1 Electron microscopy

Following sorting and centrifugation, the pellet was post-fixed with 1% osmium tetroxide in 0.1 M cacodylate buffer for 30 min, dehydrated in graded series of ethanol dilutions and embedded in Durcupan (Fluka). After polymerization, serial ultrathin sections from the pellet were cut on an Ultracut ultramicrotome (Reichert-Jung) and collected on Nickel 200 mesh square grids. Sections were counterstained with lead citrate and examined with JEM1200EX electron microscope (Jeol).

4.4.3.2 Western Blot

1.5 μg each of protein from the total cell lysate, from the 1 500 x g supernatant and from the granule extracts A and B were separated on a 10% SDS-PAGE and analyzed by Western Blot. Anti-GFP, anti-ERp29 and anti-PIST (1:2500) and anti-Rabbit IgG-HRP (1:2500) were used to reveal the presence of the subcellular markers using enhanced chemiluminescence (GE-Healthcare).

4.4.4 Granule extraction and mass spectrometry

4.4.4.1 Granule extraction

The pelleted granules were resuspended in a hypoosmotic solution composed of 10 mM EDTA in water and kept for 5 min at 4ºC. Following incubation, the pellet was extracted by three consecutive Freeze (2 min) – Boil (1 min) – Sonicate (5 sec) cycles. Following removal of the aqueous solution (extract A), the tube is rinsed with an aqueous

solution containing 10 mM EDTA and 0.025% (w/v) SDS and the content submitted to two consecutive Freeze-Boil-Sonicate cycles (extract B). In order to maximize the number and significance of the identified proteins, both successive extracts were submitted to mass spectrometry. The SDS present in the second extract was removed using the SDSAway™ sample preparation kit (Protea Biosciences) according to the manufacturer's protocol. Extracts from three purification experiments were pooled together and dried under *vacuo*.

4.4.4.2 LC-MS/MS

Protein extracts were reconstituted in 6 M urea, reduced with 1, 4-dithio-DL-threitol and alkylated with iodoacetamide. Proteolytic digestions were performed at 37ºC for 5 hrs using a sequencing-grade modified Trypsin (Promega) with a protein-enzyme ratio of 1:25. Chromatographic separation and subsequent analysis were accomplished on a nanoLC system (Eksigent) coupled to a LTQ Orbitrap hybrid mass spectrometer (ThermoFisher). The tryptic peptides were separated using self-pack PicoFrit capillary columns (75 µm i.d. x 10 cm, 15 µm tip) (New Objective) packed with Jupiter C18 reversed-phase stationary phase of 5µm particle size (Phenomenex). A gradient elution of 4-80 % acetonitrile-water (0.2 % formic acid) in 35 min was used for all separations.

4.4.4.3 Data analysis

Protein identification was performed with the Mascot software package (Matrix Science, London, UK)[26]. The search criteria were as follow: Tryptic digestion; Variable modifications include carbamidomethylation (Cys), di-methylation (Lys), di-methylation (Arg), oxidation (Met) with a peptide mass tolerance of ± 10 ppm and a fragment mass

tolerance of ± 0.7 Da. The maximum missed cleavage number was set at 2 and peptides with a score of 10 or less were excluded, while the presence of at least one bold red (high significance) peptide was required for each protein. The quality of the MS spectra was manually checked. Proteins identified as ‘‘unknown’’ were usually easily identified by searching the peptide sequence using BLAST algorithm[27] for short, nearly exact matches. In order to be included in our list, a protein must exhibit a minimum score of 50 and be identified by at least two distinct, non-overlapping peptides.

4.5 RESULTS

4.5.1 Cell culture and immunocytochemistry

AtT-20 is a tumor-derived, immortalized cell line of murine pituitary corticotropes and a recognized and accepted model for pro-opiomelanocortin (POMC) production and secretion studies. Indeed, as early as 1981, the secretory granules of these cells were isolated, studied and shown to contain mature forms corticotropin and β-lipotropin[28]. For our method to succeed, one has to label the organelle with a fluorescent molecule. In the past, green fluorescent protein (GFP) was shown to be routed to the regulated pathway of secretion and ultimately into secretory granules through linking either to a signal sequence[29] or to the NH_2-terminal domain (known as PHM) of the secretory granule resident peptidyl α-amidating monooxygenase (PAM) enzyme[25]**.** An expression vector, PHM-mGFP (the m refers to the monomeric GFP variant resulting from a A206K mutation[30]), was used to transfect AtT-20 cells and stably expressing cells were obtained. Additional studies[25] and immunocytochemistry experiments confirmed that expression of

the fusion protein, although not solely confined to secretory granules, nevertheless is limited to the regulated pathway of secretion. For example (Figure 4.1), the PHM-mGFP signal is present in secretory granules which share a pattern of distribution within the cells, of storage and of release in response to stimulation similar to adrenocorticotropin (ACTH), a hormone resulting from POMC processing.

4.5.2 Cell lysis and fluorescence-assisted organelle sorting

Potter homogenization is the most widely employed method of cell lysis though the associated mechanical stress involved is known to damage organelles and to promote formation of microsomes. We have replaced this procedure by a pre-treatment of the cells with latrunculin B, an actin filament disruptor[31] and nocodazole, a microtubule disruptor[32] in order to inhibit the secretory processes and partially dissolve the cytoskeleton to free the granules from the actin and tubulin meshes within the cells. After treatment, the cells were lysed by one freeze-thaw cycle in an isoosmotic sucrose solution which led to over 90% cell lysis. This gentle method proved highly advantageous and much less disruptive than the use of pestle strokes.

The sample was sorted using a regular cell sorter with the sample pressure lowered in an effort to minimize the shear and tear forces to which the granules are exposed. In addition to reducing the electronic background noise, it also considerably improves the resolution in the small particle size range. A narrow sorting window was selected on the density plot illustrating the fluorescence (X-axis) and size (Y-axis) of the particles in the sample. Commercially available calibration beads (non-fluorescent and fluorescent, 450 nm, 530 nm and 840 nm) were used as reference points to estimate the size and level of fluorescence of the sorted populations. As shown in Figure 4.2A, the

sorted granule population (black square) was 10 to 50 times more fluorescent than the non fluorescent beads and non-fluorescent events, while being smaller than 500 nm in size. This figure also shows all detectable events in the 1 500 x g supernatant (Figure 4.2B) and the total cell lysate (Figure 4.2D) with the trigger adjusted to minimum threshold (maximum sensitivity) on the side scatter (SSC) detector. It also illustrates how a brief centrifugation step is necessary to reveal a population (Figure 4.2A) that was not clearly distinguishable in the total cell lysate (Figure 4.2C). On average, our strict selection parameters (sorting window) allowed sorting of 3-5 million events per cell dish. Evidently, this number can vary depending on the sorting criteria; herein, it corresponds to approximately 20% of detectable events at the optimized sensitivity threshold, or 2-3% of all detectable events at maximum sensitivity.

4.5.3 Electron microscopy and Western Blot

The sorted events were pelleted, fixed and analyzed by electron microscopy. The content of a representative cut through the pelleted total lysate (Figure 4.3A) and through the FAOS-purified granules pellet (Figure. 4.3B) is shown in Figure 4.3. It is noteworthy that the granules collected are very homogeneous in size and present an intact morphology. However, most granules appear slightly decondensed, an effect likely resulting from the presence of EDTA in the sorting medium and the passage through the sorter's lines and nozzle. Samples from the total extract, the 1 500 x g supernatant and the sorted granules were routinely analyzed by SDS-PAGE and Western blotting using a battery of antibodies to assess the enrichment and depletion of specific subcellular markers. As shown in Figure 4.4, the FAOS purification leads to a considerable enrichment of the sample in mGFP (granules) with a concomitant depletion of the sample

in ERp29 (ER)[33] and PIST (Golgi)[34]. This demonstrates the efficiency of the procedure to selectively sort the granules from the rest of the mGFP-tagged material in the secretory pathway, a conclusion that is further confirmed by the mass spectrometry data.

4.5.4 Granule extraction and mass spectrometry

The purified granules were centrifuged, washed and sequentially extracted in water containing 10mM EDTA, and subsequently in water containing 0.025% SDS (w/v). The first extract was submitted to tryptic digestion and the resulting peptides were separated by liquid chromatography and analyzed with an ESI-LTQ-Orbitrap mass spectrometer. The second extract was treated (see methods) to remove SDS and processed identically. The resulting data was analyzed using the Mascot software[26]. The FAOS-purified material originating from ca. 15 million cells was routinely sufficient to identify over 100 proteins from over 3 000 peptide queries using relatively stringent criteria. Chromogranin A, a recognized secretory granule marker[35], was always and consistently the top scoring protein. Even if reproducibility between samples was excellent, data from three samples was combined to increase the number of proteins identified and their respective score. With this approach, 158 proteins were unambiguously identified. This represents more proteins than previously identified in either one of the two recent proteomic studies of endocrine secretory granule models[21, 22]. A complete list (in alphabetical order) of identified proteins is presented in Table 1 while representative MS/MS spectra are presented in Figure 4.5.

Many other important makers of the secretory granules were identified such as, for example, chromogranin B, POMC, secretogranin III and prohormone convertase PC1/3. A distribution of the identified proteins is presented in Figure 4.6 with respect to

their known, proposed or putative subcellular localizations. Overall, 27 proteins previously seen in the two above-mentioned studies[21, 22] were also identified herein. Interestingly, we also identified 27 other proteins known from the literature to be present in or associated with the secretory granules but which were never before reported in this type of organellar proteomics studies. Furthermore, 18 cytoskeleton proteins involved in transport, docking, membrane fusion and actin/tubulin network remodeling are present in our sample. These proteins are possibly associated with the granules in a transient manner, confirming that our rapid and gentle purification method leads to the identification of candidates that might not be otherwise retained during classical procedures. We identified only 5 proteins reported as being ER and Golgi residents, the only other compartments made fluorescent by the transit of our mGFP construct. This highlights the efficiency of our purification approach not only to discriminate between fluorescent and non-fluorescent organelles, but also between fluorescent compartments. Also, only 5 proteins in our list are thought to be restricted to endosomes, mitochondria, lysosomes, or proteasome. Again, this result compares advantageously to those obtained using more conventional methodology. Interestingly, we also identified proteins whose expression and localization is deregulated during metastatic transformation and that are used as tumor markers since they are found in the blood of patients (i.e. M2-pyruvate kinase[36]). If indeed present in secretory granules, such misrouting and misregulation in endocrine and neuroendocrine tumors clearly warrants further investigations.

As previously mentioned, replacing the freeze-thaw cycle for initial cell lysis by the most typically employed Potter homogenization, while keeping other steps in the protocol unchanged, led repeatedly to an important loss of information. Thus, for example, chromogranin A was no longer the top scoring protein, some secretory granule

markers were lost and serious contamination, especially by proteins from the nucleus and the mitochondria compartments, was observed (data not shown). In the same vein, relying on a gel-free approach leads to two important improvements. Firstly, we minimized sample loss frequently associated with gel staining and band extraction while allowing high quality MS data in terms of number and score of peptides from a relatively small sample as illustrated by representative MS/MS spectra (Figure 4.5). Secondly, it permitted identification of small molecular weight proteins important to granule function such as complexin 2, phosphatidylethanolamine binding protein and macrophage migration inhibitory factor, which were never previously reported in such proteomic studies

4.6 DISCUSSION

Density-based separation techniques have been the tool of choice for decades in cell fractionation experiments. Being readily accessible, equilibrium gradient centrifugation has proven quite effective in enriching a sample for a specific subcellular compartment. However, this method is not without its problems. Indeed, it usually requires large amounts of starting biological material, either from fresh tissues or from large scale cell cultures already limiting the experimenter to the study of models for which enough material is available. Also, the procedure is very resource and time-consuming, typically requiring between 48 to 72 hours of centrifugation and electrophoresis. Not only does this delay the time at which mass spectrometry data can be obtained, but it also generates concerns on how complete and accurate is the

representation from the proteins extracted from the final gel of the *in vivo* state of the organelle.

In order to circumvent this problem, we propose an alternative approach to perform organellar proteomics. Hence, we found that dissolution of the cytoskeleton followed by a freeze-thaw cycle is just as effective as pestle strokes in inducing cell lysis. This was clearly seen from the resulting MS data suggesting significant decrease in cross-contamination due to organelle breakage and leakage. Then, FAOS performed with a cell sorter was efficient in sorting out an abundant, intact and homogeneous population of dense-core secretory granules within one hour. The choice of the appropriate sorting window was rapidly established using readily available calibration beads. What is especially remarkable is the fact that even if some PHM-mGFP fluorescence is localized throughout the regulated secretory pathway (data not shown), we were nevertheless able to isolate a much enriched population of mature secretory granules. It is noteworthy that sorting a very fragile and hypoosmosis-sensitive organelle using a single fluorescent label transiting through multiple subcellular compartments was a very demanding task. This method could benefit from using recent multiple simultaneous laser wavelengths cell sorters and/or by combining constructs, fluorescent dyes and/or antibodies. Doing so, the user can virtually sort and actively exclude any event in a defined window which, in turn, leads to much increased purity of the sorted material. It would be even possible to fractionate an organelle population into specific groups depending on their state at the time of cell lysis as determined, for example, by the presence or absence of a marker. Herein, we optimized the material to be sorted out by selecting stable transfectant but we consider that our method could be of use with transient transfectant as well as with any commercially available organelle-specific dyes or conjugated antibodies.

Even in its simplest form, this method has allowed us to obtain a purified sample very rich in dense-core granules from which more than 150 proteins were identified by mass spectrometry. Many important markers (Table 1) have been identified by numerous high scoring peptides, and some important small molecular weight proteins were identified for the first time in a large scale proteomics study. Hence, one ought to expect that the isolated granules contain the appropriate peptides in good amounts. In all studies including ours, this has been the case as can be seen with proANF for atrial granules[19], insulin(s) in β-cell granules[22], neuropeptide Y, enkephalins and adrenomedullin in chromaffin granules[21], various zymogens in pancreatic zymogen granules[20] and corticotropin and proopiomelanocortin in corticotropes (this study). Associated closely with these processing products, all these studies also revealed the presence of the obligatory processing machinery including appropriate convertases, mostly PC1/3 ([21, 22], this study) and PC2 ([21, 22]), carboxypeptidase-E[19, 21, 22], and PAM ([19, 21, 22] and this study). It is noteworthy, however, that these enzymes do not appear to be routinely observed as, for example, we were not able to detect CpE nor was the study on chromaffin granules able to detect PAM[21]. Similarly, various recognized granules peptides such as chromogranin A ([21, 22] and this study) and B ([19] and this study), a variable mixture of secretogranins ([21, 22] and this study), 7B2[21, 22], proSAAS[21, 22] or granule membrane proteins such as cytB561 ([21]) and V-ATPase ([19, 21, 22] and this study) were reported. All the granules preparations contained characteristic but distinct set of Rab proteins and VAMPs though none of the latter was seen in the present study nor was there any VAMP4 detected in any study. Interestingly, VAMP4 together with a great many number of SNARE and Rab proteins were shown to be abundant in synaptosomes[23, 24].

Furthermore, proteins reported to be associated with secretory granules were observed; these include kalirin[19], myosins ([20] and this study), tubulin and actin ([21] and this study). Actually, prior to our inclusion of actin/tubulin network disruptors, we observed a much larger distribution of fluorescent events likely due to aggregation and/or tethering of granules to actin/tubulin as previously reported[37]. Hence, our observation is much in favor with their close association with the granules as envisioned. Such proteins involved in these transient and dynamic interactions would not necessarily have been identified if the purification steps had required more than a few hours. Moreover, it is certainly worth noting that each study possesses its own set of unique proteins whose presence and functional role remain to be explored and established. In our study, some of these include the 14-3-3 protein, the Na/K-ATPase, MARCKS, various chaperonins, secretory signal-peptide devoid proteins such as cyclophilin A, serpin-1 and stathmin. The only major unforeseen proteins originated from ribosomes. However, we cannot rule out the possibility that some PHM-mGFP bound to polyribosomes could have been co-sorted during FAOS. Although ribosomal proteins are present in our list, their significance must also be kept in perspective. Effectively, as a group, they are identified by an average of 3 peptides and a score of 109, which is well below chromogranin A with its 26 peptides for a score of 1440. Moreover, it is difficult to compare this type of contamination with that of the other two proteomic studies which, by design, would probably not identify most of these proteins as they are very small and cannot be effectively resolved by gel electrophoresis.

In conclusion, we present here a very rapid, simple and adaptable method to purify organelles and obtain their proteome. Requiring relatively low amounts of starting materials, the protocol can be performed within a day. In this presented form, the method

might not immediately replace the conventional protocols for subcellular fractionation, but the availability of hundreds of fluorescent molecules, including dyes and antibodies, combined with the speed, power and versatility of flow cytometry could rapidly contribute to overcoming some of the challenges related to sample preparation in organellar proteomics.

4.7 ACKNOWLEDGMENTS

We thank Dany Gauthier for technical expertise and Nadia Rabah for help in cell culture, Annie Vallée for processing the samples for electron microscopy and Denis Faubert for LC-MS/MS analyzes (IRCM). Special thanks to Eric Massicotte and Martine Dupuis for their precious help in optimizing and running the flow cytometry experiments in the IRCM Cytometry Unit. This work was supported by a Canadian Institutes of Health Research operating grant to CL (MT-74479) and by National Institutes of Health grant to REM (DK-32948). D.J. Gauthier is a recipient of a Fonds de la recherche en santé du Québec (FRSQ) studentship, and J.A. Sobota holds a National Research Service Award (DE-017094).

4.8 FIGURES AND LEGENDS

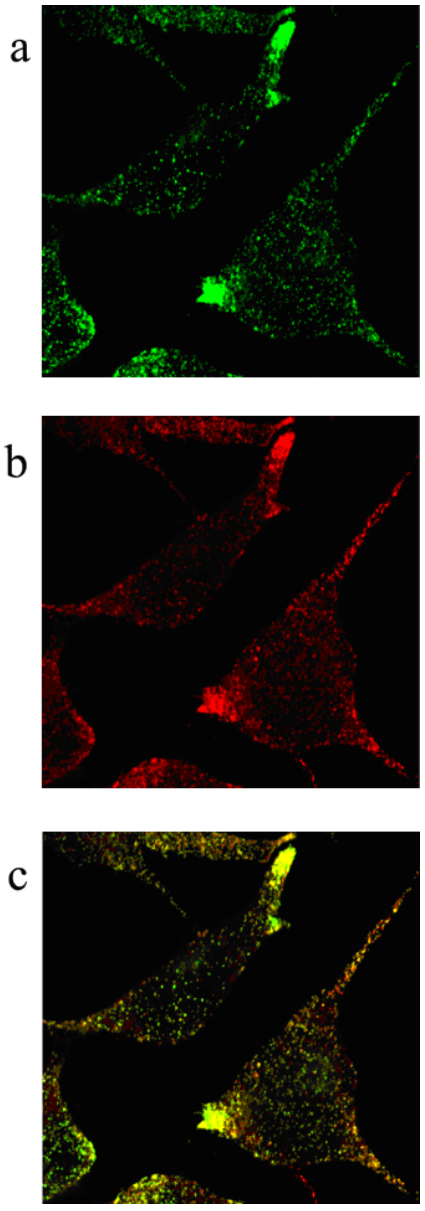


Figure 4.1: Intracellular localization of the PHM-mGFP fusion protein within AtT-20 transfected cells. (A) PHM-mGFP (green) (B) ACTH (red) and (C) merging of colors highlights the colocalization of PHM-mGFP and ACTH in the secretory granules at the tips of the cells.

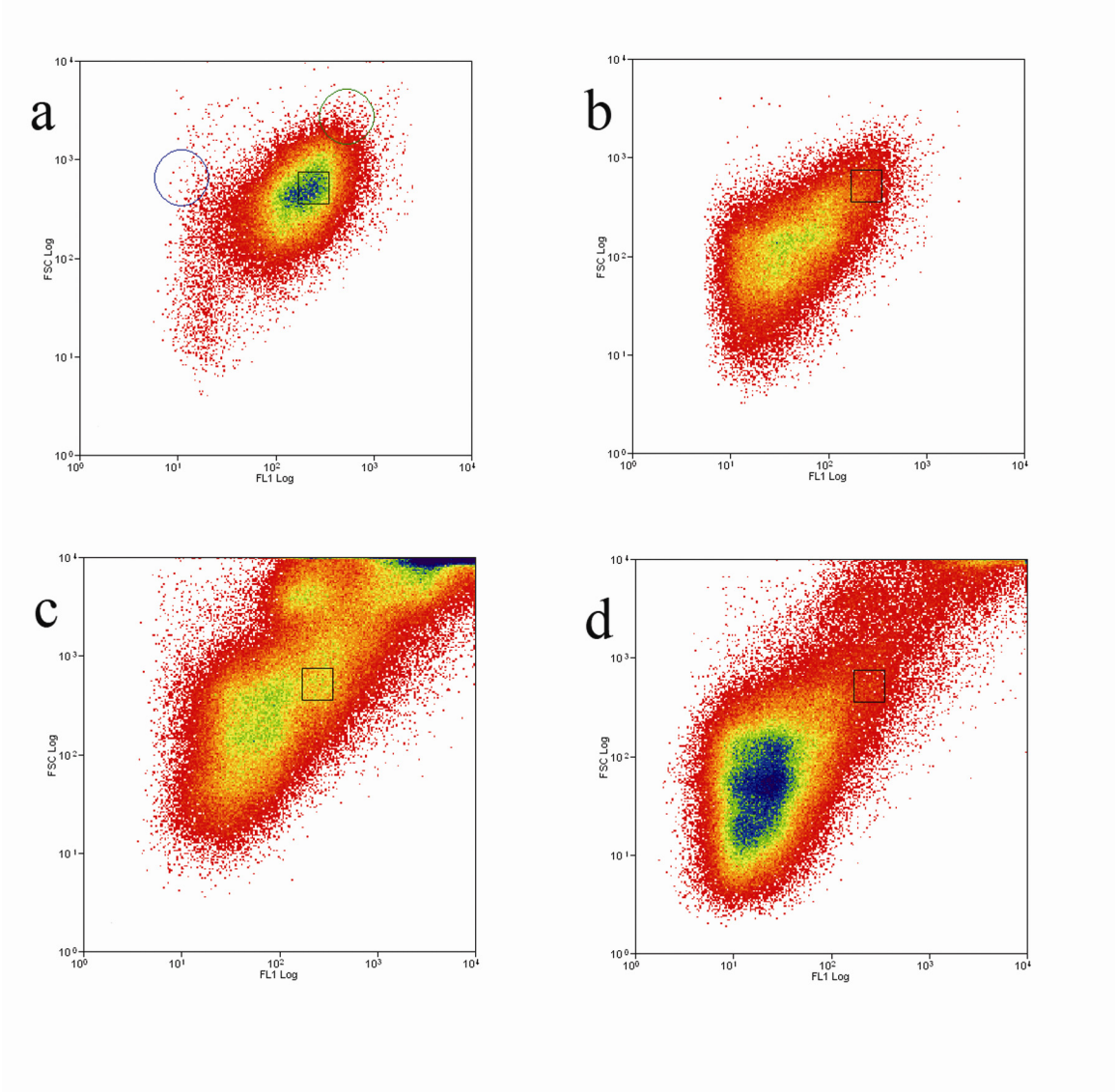


Figure 4.2*:* Fluorescence-assisted organelle sorting of the secretory granules (FAOS). (A) to (D are density plots (green fluorescence (X-axis) versus size (Y-axis) of all events detected by the cell sorter under each set of conditions. The black square represents the selected sorting region which corresponds to the secretory granules (see text). (A) 1 500 x g supernatant analysis with the threshold set to optimize the detection of the secretory granule population. The blue circle and green circle indicate the region in which non-fluorescent 450 nm beads and fluorescent 840 nm beads are respectively

located when analyzed. (B) 1 500 x g supernatant analysis with the threshold set to the minimum (maximum sensitivity), allowing visualization of the most abundant smaller, low-fluorescent to non-fluorescent events. (C) Total cell lysate analyzed as in (A). The presence of numerous cellular particles of varying fluorescence is visible, masking the granule population. The top right density represents larger fluorescent particles such as intact cells and large ER and Golgi debris. (D) Total cell lysate analyzed as in (B). In order to present informative and representative plots, 75 000 events were pooled in (A) and (B) while 225 000 events were pooled in (C) and (D). Density is color-coded from low (few events) in red to high (many events) in blue.

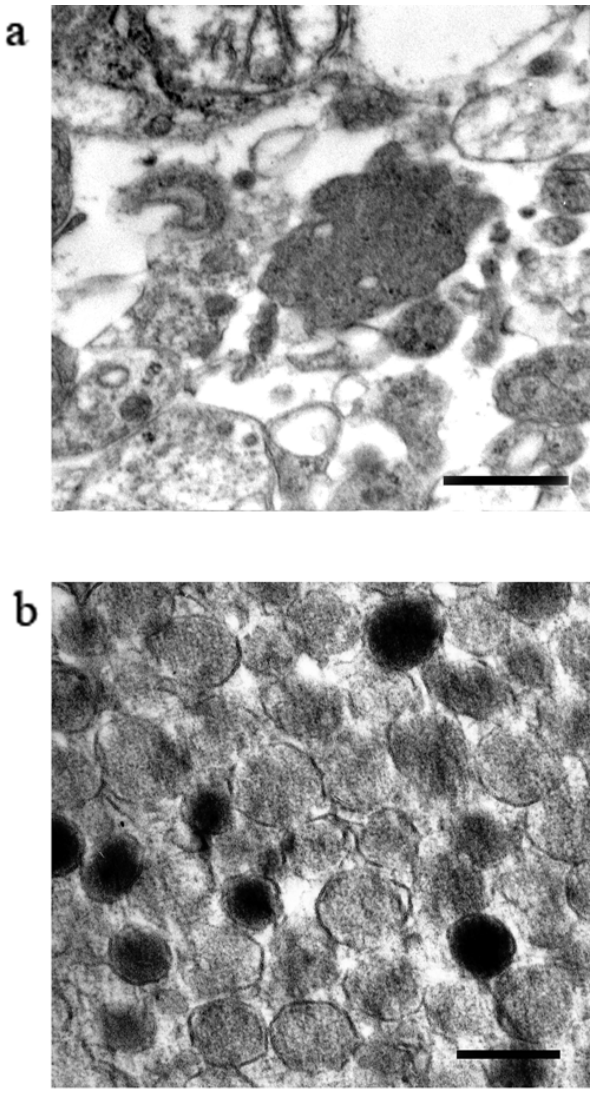


Figure 4.3: Secretory granules visualized by electron microscopy. (A) A representative cut of the total cell extract; bar: 500 nm. (B) A representative cut thought the purified secretory granules pellet shows a population of intact, membrane bound granules, homogeneous in size. Most of them are slightly decondensed (see text); bar: 200 nm.

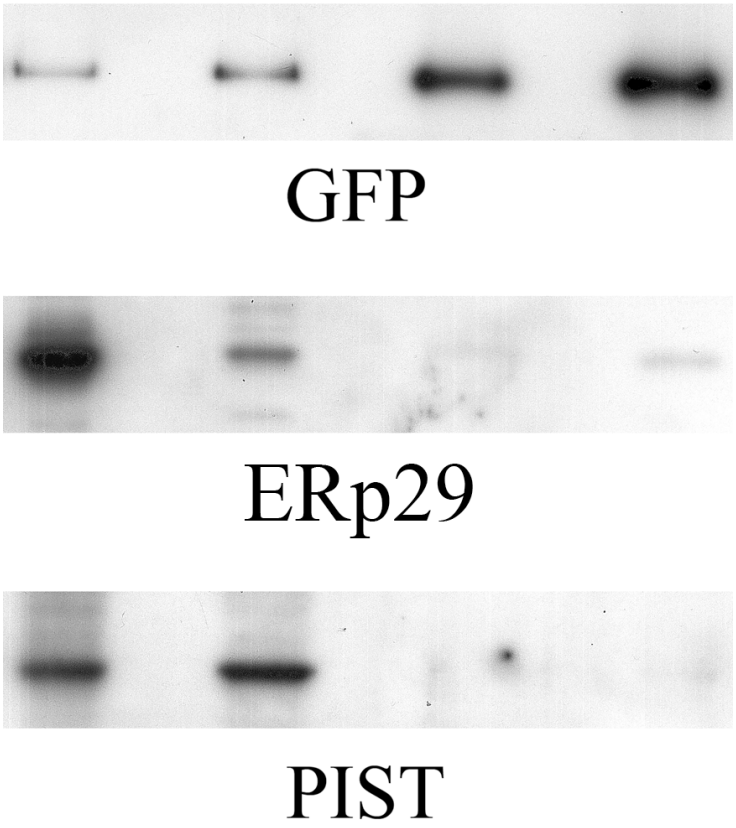
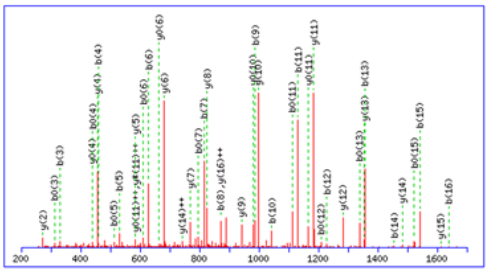
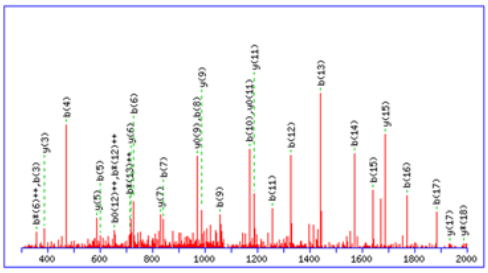


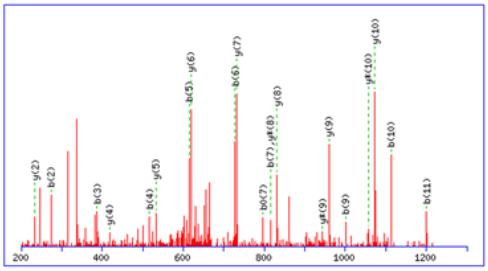
Figure 4.4: **Western blots illustrating the granules enrichment.** From left to right, separated by empty lanes are the total cell lysate, the 1 500 x g supernatant, the granule extract A and the granule extract B. 1.5µg of protein was loaded in each lane. (A) Anti-GFP antibody. (B) Anti-ERp29 antibody (ER). (C) Anti-PIST antibody (Golgi). While the ER and Golgi markers are depleted, the GFP signal, present only in the ER, Golgi and secretory granules, is enriched.



AEEEAVWGDGSPEPSPR
POMC peptide, score 112



AEDQELESLSAIEAELEK
Chromogranin A peptide, score 137



CIIEVLSNALSK
Chromogranin B peptide, score 71

Figure 4.5: Examples of representative MS/MS spectra from secretory granule markers.

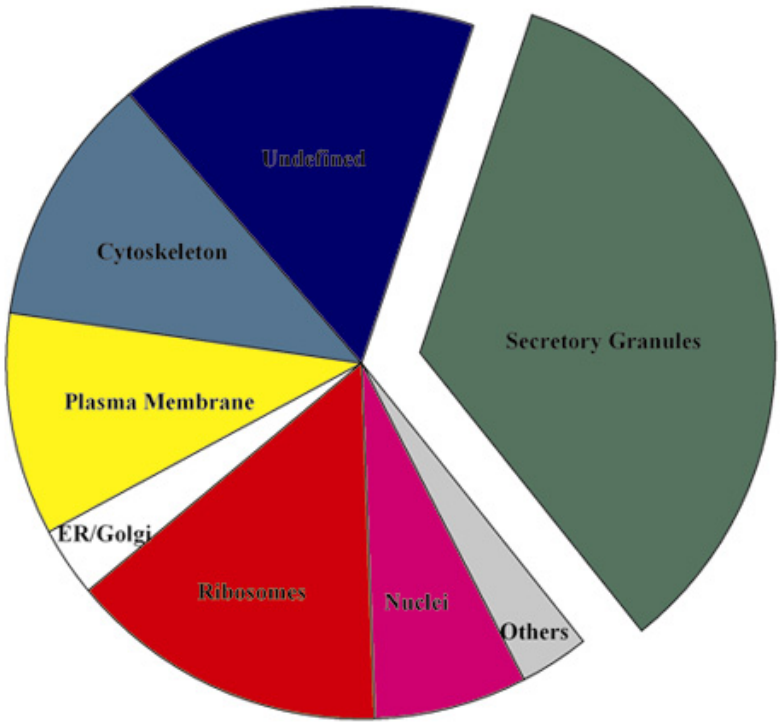


Figure 4.6: Pie chart illustrating the distribution of the identified proteins with respect to their known, proposed or putative subcellular localizations. Pie slices correspond to Cytoskeleton (structure, trafficking, fusion); Granules (protein identified in the literature as being part of or associated with the secretory granules); Nucleus; Other (Endosome, Lysosome, Mitochondria, Proteasome); Plasma membrane; Ribosome; ER/Golgi; Undefined (protein present in more than one major location, or for which there is not enough information available).

4.9 TABLE AND LEGEND

Table 4.1

Name	Accession	Peptides	Score	Category
14-3-3 protein epsilon	gi|74189349	8	408	Granule Literature[38]
14-3-3 protein gamma	gi|3065929	7	351	Granule Proteomics
14-3-3 zeta	gi|1841387	11	576	Granule Proteomics
78 kDa glucose-regulated protein	gi|1304157	9	480	Secretory Pathway
A kinase anchor protein (PRKA)	gi|39930557	7	106	Variable
Acrogranin / Progranulin	gi|50852	4	230	Variable
Actin, gamma	gi|74213524	21	1107	Granule Proteomics
Actinin, alpha 1	gi|32766260	2	129	Granule Literature[39]
Adenomatous polyposis coli protein	gi|12643510	3	50	Cytoskeleton
Alpha-NAC	gi|1666692	4	252	Nucleus
Ankyrin repeat-containing cofactor 1	gi|82919283	3	52	Variable
Asparagine synthetase B	gi|26345208	3	94	Variable
ATP synthase, beta subunit	gi|23272966	4	136	Granule Proteomics
ATPase, Na/K transporting, alpha 1 polypeptide	gi|21595127	23	1115	Plasma Membrane
ATPase, Na/K transporting, beta subunit	gi|54130	4	156	Plasma Membrane
Calmodulin	gi|74143933	8	573	Granule Literature[40]
Calreticulin	gi|26344403	3	107	Secretory Pathway
CAP, adenylate cyclase-associated protein 1	gi|26343401	2	117	Cytoskeleton
Carnitine deficiency-associated gene 3B	gi|28461294	5	203	Variable
CAST1/ERC2	gi|38231910	3	54	Granule Literature[41]
Catenin (Ctnnd1) p120	gi|26006157	3	152	Plasma Membrane
Cellular nucleic acid binding protein	gi|50471	2	85	Nucleus
Chaperonin containing TCP-1 delta subunit	gi|460317	2	83	Granule Literature[42]
Chaperonin containing TCP-1 theta subunit	gi|5295992	4	91	Granule Literature[42]

Chaperonin containing TCP-1, beta subunit	gi|468546	3	97	Granule Literature[42]
Chaperonin subunit 5	gi|37359776	4	243	Variable
Chaperonin subunit 6a	gi|17391080	2	79	Variable
Charged multivesicular body protein (CHMP4b)	gi|74192743	4	117	Other Organelle
Chromogranin A	gi|20071660	26	1440	Granule Proteomics
Chromogranin B	gi|50409	18	887	Granule Proteomics
Clathrin, heavy polypeptide	gi|33438248	6	233	Granule Proteomics
Coatomer protein complex, gamma subunit	gi|19354315	2	58	Secretory Pathway
Cofilin 1	gi|12861068	2	64	Cytoskeleton
Complexin 2	gi|62740137	2	84	Granule Literature[43]
Cp27	gi|3115274	2	56	Plasma Membrane
Cu/Zn superoxide dismutase	gi|226471	8	489	Granule Proteomics
Cyclophilin A (Peptidyl-prolyl cis-trans isomerase A)	gi|74146841	9	407	Granule Proteomics
Cysteine-rich protein 1	gi|74212172	4	127	Cytoskeleton
DEAD box polypeptide 3	gi|18204785	3	147	Nucleus
Desmoplakin	gi|82950147	5	111	Cytoskeleton
Diazepam binding inhibitor	gi|67511482	2	110	Granule Literature[44]
Dynein, heavy chain 8	gi|14335452	4	53	Granule Literature[45]
Dynein, heavy chain 1	gi|13384736	4	70	Granule Literature[45]
Dynein, light chain roadblock-type 1	gi|12842877	2	56	Cytoskeleton
EEF2	gi|26324898	5	165	Nucleus
EIF4a1	gi|50820	4	146	Nucleus
EIF4h	gi|39104480	4	121	Nucleus
Elongation factor Tu	gi|26345590	17	819	Nucleus
Enolase 1, alpha	gi|53734652	16	1030	Variable
Fatty acid synthase	gi|74142919	3	129	Plasma Membrane
Fatty acid-binding protein, epidermal	gi|74191310	6	246	Granule Literature[46]

Fructose-bisphosphate aldolase A	gi|7548322	17	687	Granule Proteomics
Gag-pro-pol polyprotein	gi|559059	8	245	Variable
Gasdermin 1	gi|26325036	3	102	Variable
Glucose phosphate isomerase	gi|56554102	7	309	Granule Literature[47]
Glyceraldehyde-3-phosphate dehydrogenase	gi|62201487	14	725	Granule Proteomics
Glycyl-tRNA synthetase	gi|18256043	2	55	Variable
GP42/Basigin protein	gi|631683	5	212	Plasma Membrane
Guanine nucleotide binding protein, alpha subunit, inhibitory	gi|309255	2	83	Granule Literature[48]
Guanine nucleotide binding protein, alpha subunit, stimulating	gi|47271396	7	232	Granule Proteomics
Guanine nucleotide binding protein, beta 2 subunit	gi|984551	7	232	Granule Proteomics
Heat shock protein hsp70	gi|74220592	25	1181	Granule Literature[49]
Heat-shock protein hsp84/90	gi|74147026	29	1330	Variable
Heterogeneous nuclear ribonucleoprotein K	gi|12847547	2	52	Variable
HN1	gi|74220270	2	125	Variable
HN1-like protein	gi|74213941	4	192	Variable
Importin beta 1	gi|2829480	3	144	Variable
Integrin alpha 6	gi|110577	3	58	Plasma Membrane
IQ motif containing GTPase activating protein 1	gi|27370648	3	70	Cytoskeleton
Lactate dehydrogenase 1	gi|74198692	13	636	Variable
L-type amino acid transporter 1	gi|6906727	2	145	Plasma Membrane
M2-type pyruvate kinase	gi|74221210	23	1389	Variable
Macrophage migration inhibitory factor, C	gi|5822094	3	69	Granule Literature[50]
Microtubule-actin crosslinking factor 1	gi|94372354	16	162	Cytoskeleton
Monocarboxylate transporter 4	gi|7688756	2	54	Plasma Membrane
Myosin, heavy polypeptide	gi|20137006	19	705	Granule Proteomics
Myosin, light polypeptide	gi|16924329	6	345	Granule Literature[51]
Myosin, regulatory light chain 2	gi|12851268	3	121	Granule Literature[51]

Myosin-I, beta	gi|1924961	2	88	Cytoskeleton
Myristoylated alanine rich protein kinase C substrate (MARCKS)	gi|6678768	2	71	Plasma Membrane
Nebulin	gi|94366489	14	126	Cytoskeleton
Neural cell adhesion molecule 1 (NCAM)	gi|817984	10	480	Granule Literature[52]
Nucleolin	gi|13529464	6	168	Nucleus
Nucleosome assembly protein 1-like 1	gi|50722	3	109	Nucleus
Oxidation resistance protein 1	gi|58047714	3	51	Plasma Membrane
P23	gi|12843224	2	122	Variable
PDGFA associated protein	gi|74151229	5	153	Variable
Peptidylglycine alpha-amidating monooxygenase	gi|7305367	4	193	Granule Proteomics
Peroxiredoxin 1	gi|12846314	3	67	Variable
Phosphatidylethanolamine binding protein	gi|1517864	4	113	Granule Literature[53]
Phosphofructokinase-1 C isozyme	gi|74224916	4	82	Plasma Membrane
Phosphoglycerate kinase 1	gi|129903	7	438	Variable
Phosphoglycerate mutase 1	gi|12844989	3	95	Variable
Plakoglobin	gi|423532	2	102	Cytoskeleton
Profilin 1	gi|26389590	2	64	Cytoskeleton
Prolyl 4-hydroxylase	gi|54777	8	163	Secretory Pathway
Proopiomelanocortin (POMC)	gi|74227361	9	517	Granule Literature [54]
Proprotein convertase PC1/3	gi|50055	2	79	Granule Proteomics
Proteasome, 26S non-ATPase subunit 2	gi|12861131	2	96	Other Organelle
Proteasome, delta subunit	gi|74191020	2	66	Other Organelle
Protein disulfide isomerase associated 3	gi|6679687	8	247	Granule Proteomics
Rab10	gi|10435058	2	112	Granule Proteomics
Rab14	gi|10435483	2	70	Granule Proteomics
Rab1A	gi|74147521	2	109	Granule Proteomics
Radixin	gi|74186081	6	106	Cytoskeleton

Rap1a	gi|53236951	3	67	Granule Literature[55]
Regulator G-protein signaling 9	gi|2739458	4	53	Plasma Membrane
Ribosomal protein L10A	gi|6755350	3	86	Ribosome
Ribosomal protein L10E	gi|71051393	3	108	Ribosome
Ribosomal protein L15	gi|12846287	3	82	Ribosome
Ribosomal protein L17	gi|12832997	3	88	Ribosome
Ribosomal protein L18	gi|12840700	3	132	Ribosome
Ribosomal protein L19	gi|6677773	2	66	Ribosome
Ribosomal protein L21E	gi|94368373	4	98	Ribosome
Ribosomal protein L23	gi|12849613	2	160	Ribosome
Ribosomal protein L26	gi|74179650	3	94	Ribosome
Ribosomal protein L3	gi|74198856	3	77	Ribosome
Ribosomal protein L6	gi|695638	6	133	Ribosome
Ribosomal protein L8	gi|74203516	3	77	Ribosome
Ribosomal protein P2	gi|74140891	5	290	Ribosome
Ribosomal protein S10	gi|74198792	2	72	Ribosome
Ribosomal protein S16	gi|7305445	3	54	Ribosome
Ribosomal protein S18	gi|198578	2	85	Ribosome
Ribosomal protein S2	gi|12835827	2	70	Ribosome
Ribosomal protein S20	gi|74181462	3	73	Ribosome
Ribosomal protein S23	gi|12846275	2	68	Ribosome
ribosomal protein S28	gi|12833257	3	140	Ribosome
Ribosomal protein S3	gi|12847921	2	57	Ribosome
Ribosomal protein S4	gi|12846200	4	128	Ribosome
Ribosomal protein S8	gi|26353710	3	138	Ribosome
Secretogranin II	gi|54096	2	79	Granule Proteomics
Secretogranin III	gi|74211039	3	147	Granule Proteomics

Serpin1	gi|12836024	3	81	Granule Literature[56]
Seryl-aminoacyl-tRNA synthetase 1	gi|26349967	3	89	Other Organelle
Similar to PEST-containing nuclear protein (PCNP)	gi|74141568	2	94	Nucleus
Similar to tumor protein, translationally-controlled 1 (TCTP)	gi|74181622	2	55	Plasma Membrane
Similar to ubiquitin C	gi|94375393	3	93	Granule Proteomics
Solute carrier family 3 (CD98)	gi|26354873	12	528	Plasma Membrane
Spectrin, beta 2	gi|7106421	2	54	Granule Literature[57]
Src substrate cortactin	gi|2498955	5	76	Cytoskeleton
Stathmin	gi|12832714	13	463	Cytoskeleton
Sulfated glycoprotein (prosaposin)	gi|34328185	3	126	Granule Proteomics
Synaptosomal-associated protein 25 (SNAP25)	gi|26346913	2	88	Granule Literature[58]
Synuclein, beta	gi|18043841	2	103	Cytoskeleton
Thymosin, beta 10	gi|74199914	2	91	Cytoskeleton
Titin isoform N2-A	gi|77812697	13	139	Cytoskeleton
Transferrin receptor 2	gi|15559221	3	52	Other Organelle
Triosephosphate isomerase	gi|54855	3	140	Variable
Tropomyosin	gi|74213492	22	959	Granule Literature[59]
Tubulin, alpha 1a	gi|74202338	17	991	Granule Proteomics
Tubulin, beta	gi|74141821	21	1381	Granule Literature[60]
Tumor metastatic process-associated protein NM23	gi|387496	2	57	Secretory Pathway
Ubiquitin carboxy-terminal hydrolase L1	gi|6755929	2	60	Plasma Membrane
Ubiquitin-activating enzyme E1	gi|74228573	3	62	Variable
Ubiquitin protein ligase	gi|94399459	8	85	Variable
Vacuolar H+ATPase B2	gi|74195936	3	53	Granule Proteomics
VcpP97 (valosin)	gi|62738728	4	94	Granule Proteomics
Y box transcription factor	gi|199821	9	491	Nucleus

Table 4.1: List of all identified proteins. The table lists the accession number, the number of peptides matched and the Mascot score of the protein. When the accession number matched to an unknown protein in the database, the identity of the protein was usually easily determined by searching, using the BLAST algorithm, the amino acid sequence of the identified peptides in the NCBI databases. The proteins were classified according to their major subcellular localization into the following categories: Cytoskeleton (structure, transport, fusion), Granule Literature (proteins known from the literature to be part of or associated to the secretory granule), Granule Proteomics (proteins identified in one or both of the organellar proteomics studies targeting the chromaffin or insulin-containing granules), Nucleus, Other Organelle (Endosome, Mitochondria, Lysosome, Proteasome), Plasma Membrane, Ribosome, Secretory Pathway, Variable (proteins having multiple major localization sites or for which not enough information is available).

4.10 BIBLIOGRAPHY

1. Yates, J.R., 3rd, Gilchrist, A., Howell, K.E. & Bergeron, J.J. Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell. Biol.* **6**, 702-714 (2005).

2. Murphy, R.F. Analysis and isolation of endocytic vesicles by flow cytometry and sorting: demonstration of three kinetically distinct compartments involved in fluid-phase endocytosis. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8523-8526 (1985).

3. Ho, E., Hayen, A. & Wilkins, M.R. Characterisation of organellar proteomes: a guide to subcellular proteomic fractionation and analysis. *Proteomics* **6**, 5746-5757 (2006).

4. Schirmer, E.C., Florens, L., Guan, T., Yates, J.R., 3rd & Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301**, 1380-1382 (2003).

5. Andersen, J.S. et al. Nucleolar proteome dynamics. *Nature* **433**, 77-83 (2005).

6. Gilchrist, A. et al. Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265-1281 (2006).

7. Takatalo, M.S., Kouvonen, P., Corthals, G., Nyman, T.A. & Ronnholm, R.H. Identification of new Golgi complex specific proteins by direct organelle proteomic analysis. *Proteomics* **6**, 3502-3508 (2006).

8. Casey, T.M., Meade, J.L. & Hewitt, E.W. Organelle proteomics: identification of the exocytic machinery associated with the natural killer cell secretory lysosome. *Mol. Cell. Proteomics* **6**, 767-780 (2007).

9. Stuart, L.M. et al. A systems biology analysis of the Drosophila phagosome. *Nature* **445**, 95-101 (2007).

10. Borner, G.H., Harbour, M., Hester, S., Lilley, K.S. & Robinson, M.S. Comparative proteomics of clathrin-coated vesicles. *J. Cell Biol.* **175**, 571-578 (2006).

11. Prokisch, H. et al. Integrative analysis of the mitochondrial proteome in yeast. *PLoS Biol.* **2**, e160 (2004).

12. Huber, L.A., Pfaller, K. & Vietor, I. Organelle proteomics: implications for subcellular fractionation in proteomics. *Circ. Res.* **92**, 962-968 (2003).

13. Givan, A.L. Flow cytometry: an introduction. *Methods Mol. Biol.* **263**, 1-32 (2004).

14. Bernas, T., Gregori, G., Asem, E.K. & Robinson, J.P. Integrating cytomics and proteomics. *Mol. Cell. Proteomics* **5**, 2-13 (2006).

15. Wilson, R.B. & Murphy, R.F. Flow-cytometric analysis of endocytic compartments. *Methods Cell Biol.* **31**, 293-317 (1989).

16. Cossarizza, A., Ceccarelli, D. & Masini, A. Functional heterogeneity of an isolated mitochondrial population revealed by cytofluorometric analysis at the single organelle level. *Exp. Cell Res.* **222**, 84-94 (1996).

17. Dhandayuthapani, S. et al. Green fluorescent protein as a marker for gene expression and cell biology of mycobacterial interactions with macrophages. *Mol. Microbiol.* **17**, 901-912 (1995).

18. Bock, G., Steinlein, P. & Huber, L.A. Cell biologists sort things out: Analysis and purification of intracellular organelles by flow cytometry. *Trends Cell Biol.* **7**, 499-503 (1997).

19. Muth, E., Driscoll, W.J., Smalstig, A., Goping, G. & Mueller, G.P. Proteomic analysis of rat atrial secretory granules: a platform for testable hypotheses. *Biochim. Biophys. Acta* **1699**, 263-275 (2004).

20. Chen, X. et al. Organellar proteomics: analysis of pancreatic zymogen granule membranes. *Mol. Cell. Proteomics* **5**, 306-312 (2006).

21. Wegrzyn, J., Lee, J., Neveu, J.M., Lane, W.S. & Hook, V. Proteomics of Neuroendocrine Secretory Vesicles Reveal Distinct Functional Systems for Biosynthesis and Exocytosis of Peptide Hormones and Neurotransmitters. *J. Proteome Res.* **6**, 1652-1665 (2007).

22. Brunner, Y. et al. Proteomic analysis of insulin secretory granules. *Mol. Cell. Proteomics* (2007).

23. Takamori, S. et al. Molecular anatomy of a trafficking organelle. *Cell* **127**, 831-846 (2006).

24. Burre, J. & Volknandt, W. The synaptic vesicle proteome. *J. Neurochem.* **101**, 1448-1462 (2007).

25. Sobota, J.A., Ferraro, F., Back, N., Eipper, B.A. & Mains, R.E. Not all secretory granules are created equal: Partitioning of soluble content proteins. *Mol. Biol. Cell* **17**, 5038-5052 (2006).

26. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).

27. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

28. Gumbiner, B. & Kelly, R.B. Secretory granules of an anterior pituitary cell line, AtT-20, contain only mature forms of corticotropin and beta-lipotropin. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 318-322 (1981).

29. El Meskini, R. et al. A signal sequence is sufficient for green fluorescent protein to be routed to regulated secretory granules. *Endocrinology* **142**, 864-873 (2001).

30. Zacharias, D.A., Violin, J.D., Newton, A.C. & Tsien, R.Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913-916 (2002).

31. de Oliveira, C.A. & Mantovani, B. Latrunculin A is a potent inhibitor of phagocytosis by macrophages. *Life Sci.* **43**, 1825-1830 (1988).

32. Vasquez, R.J., Howell, B., Yvon, A.M., Wadsworth, P. & Cassimeris, L. Nanomolar concentrations of nocodazole alter microtubule dynamic instability in vivo and in vitro. *Mol. Biol. Cell* **8**, 973-985 (1997).

33. Demmer, J., Zhou, C. & Hubbard, M.J. Molecular cloning of ERp29, a novel and widely expressed resident of the endoplasmic reticulum. *FEBS Lett.* **402**, 145-150 (1997).

34. Neudauer, C.L., Joberty, G. & Macara, I.G. PIST: a novel PDZ/coiled-coil domain binding partner for the rho-family GTPase TC10. *Biochem. Biophys. Res. Commun.* **280**, 541-547 (2001).

35. Hendy, G.N., Bevan, S., Mattei, M.G. & Mouland, A.J. Chromogranin A. *Clin. Invest. Med.* **18**, 47-65 (1995).

36. Oremek, G.M., Rox, S., Mitrou, P., Sapoutzis, N. & Sauer-Eppel, H. Tumor M2-PK levels in haematological malignancies. *Anticancer Res.* **23**, 1135-1138 (2003).

37. Abu-Hamdah, R., Cho, W.J., Horber, J.K. & Jena, B.P. Secretory vesicles in live cells are not free-floating but tethered to filamentous structures: a study using photonic force microscopy. *Ultramicroscopy* **106**, 670-673 (2006).

38. Roth, D. et al. Characterization of 14-3-3 proteins in adrenal chromaffin cells and demonstration of isoform-specific phospholipid binding. *Biochem. J.* **301 (Pt 1)**, 305-310 (1994).

39. Baldini, G. et al. Rabphilin localizes with the cell actin cytoskeleton and stimulates association of granules with F-actin cross-linked by {alpha}-actinin. *J. Biol. Chem.* **280**, 34974-34984 (2005).

40. Taylor, K.A. Regulation and recycling of myosin V. *Curr. Opin. Cell Biol.* **19**, 67-74 (2007).

41. Inoue, E. et al. ELKS, a protein structurally related to the active zone protein CAST, is involved in Ca2+-dependent exocytosis from PC12 cells. *Genes Cells* **11**, 659-672 (2006).

42. Creutz, C.E., Liou, A., Snyder, S.L., Brownawell, A. & Willison, K. Identification of the major chromaffin granule-binding protein, chromobindin A, as the cytosolic chaperonin CCT (chaperonin containing TCP-1). *J. Biol. Chem.* **269**, 32035-32038 (1994).

43. Bajohrs, M., Darios, F., Peak-Chew, S.Y. & Davletov, B. Promiscuous interaction of SNAP-25 with all plasma membrane syntaxins in a neuroendocrine cell. *Biochem. J.* **392**, 283-289 (2005).

44. Metz-Boutigue, M.H., Goumon, Y., Lugardon, K., Strub, J.M. & Aunis, D. Antibacterial peptides are present in chromaffin cell secretory granules. *Cell. Mol. Neurobiol.* **18**, 249-266 (1998).

45. Kraemer, J., Schmitz, F. & Drenckhahn, D. Cytoplasmic dynein and dynactin as likely candidates for microtubule-dependent apical targeting of pancreatic zymogen granules. *Eur. J. Cell Biol.* **78**, 265-277 (1999).

46. Yanase, H., Shimizu, H., Kanda, T., Fujii, H. & Iwanaga, T. Cellular localization of the diazepam binding inhibitor (DBI) in the gastrointestinal tract of mice and its coexistence with the fatty acid binding protein (FABP). *Arch. Histol. Cytol.* **64**, 449-460 (2001).

47. Niinaka, Y., Paku, S., Haga, A., Watanabe, H. & Raz, A. Expression and secretion of neuroleukin/phosphohexose isomerase/maturation factor as autocrine motility factor by tumor cells. *Cancer Res.* **58**, 2667-2674 (1998).

48. Sattar, A.A., Boinpally, R., Stromer, M.H. & Jena, B.P. G(alpha)(i3) in pancreatic zymogen granules participates in vesicular fusion. *J. Biochem. (Tokyo).* **131**, 815-820 (2002).

49. Evdonin, A.L. et al. The release of Hsp70 from A431 carcinoma cells is mediated by secretory-like granules. *Eur. J. Cell Biol.* **85**, 443-455 (2006).

50. Nishino, T. et al. Localization of macrophage migration inhibitory factor (MIF) to secretory granules within the corticotrophic and thyrotrophic cells of the pituitary gland. *Mol. Med.* **1**, 781-788 (1995).

51. Neco, P. et al. New roles of myosin II during vesicle transport and fusion in chromaffin cells. *J. Biol. Chem.* **279**, 27450-27457 (2004).

52. Chan, S.A., Polo-Parada, L., Landmesser, L.T. & Smith, C. Adrenal chromaffin cells exhibit impaired granule trafficking in NCAM knockout mice. *J. Neurophysiol.* **94**, 1037-1047 (2005).

53. Goumon, Y. et al. Identification of morphine-6-glucuronide in chromaffin cell secretory granules. *J. Biol. Chem.* **281**, 8082-8089 (2006).

54. Loh, Y.P., Kim, T., Rodriguez, Y.M. & Cawley, N.X. Secretory granule biogenesis and neuropeptide sorting to the regulated secretory pathway in neuroendocrine cells. *J. Mol. Neurosci.* **22**, 63-71 (2004).

55. Fukai, S., Matern, H.T., Jagath, J.R., Scheller, R.H. & Brunger, A.T. Structural basis of the interaction between RalA and Sec5, a subunit of the sec6/8 complex. *EMBO J.* **22**, 3267-3278 (2003).

56. Hook, V. & Metz-Boutigue, M.H. Protein trafficking to chromaffin granules and proteolytic processing within regulated secretory vesicles of neuroendocrine chromaffin cells. *Ann. N. Y. Acad. Sci.* **971**, 397-405 (2002).

57. Perrin, D. & Aunis, D. Reorganization of alpha-fodrin induced by stimulation in secretory cells. *Nature* **315**, 589-592 (1985).

58. Chieregatti, E., Witkin, J.W. & Baldini, G. SNAP-25 and synaptotagmin 1 function in Ca2+-dependent reversible docking of granules to the plasma membrane. *Traffic* **3**, 496-511 (2002).

59. Lin, J.J., Warren, K.S., Wamboldt, D.D., Wang, T. & Lin, J.L. Tropomyosin isoforms in nonmuscle cells. *Int. Rev. Cytol.* **170**, 1-38 (1997).

60. Baltrusch, S. & Lenzen, S. Novel Insights Into the Regulation of the Bound and Diffusible Glucokinase in MIN6 {beta}-Cells. *Diabetes* **56**, 1305-1315 (2007).

Chapter 5

General discussion and conclusions

5.1 SUBCELLULAR PROTEOMICS

As described in the introduction, the field of subcellular proteomics has been quite successful in decreasing the complexity of the samples and data originating from whole cell or tissue proteomes. It is increasingly understood that the most biologically relevant information in proteomics can be obtained if the proteins identified can interact with each other and/or are part of the same microenviroment. Whereas this information was lacking in the early and mostly descriptive proteomic analyzes of whole cell extracts, some of it can now be derived from subcellular proteomics studies, highlighting the power of this field in the study of small but complex functional structures such as organelles. The elucidation of several genomes and the evolution of the tools and techniques employed, especially as far as protein identification and data analysis are concerned, now allow the researchers to obtain organellar proteomes that are probably close to completeness. One of the obvious practical applications for this is therefore the identification of novel biomarkers for diseases in terms of diagnosis, therapeutic and prognostic purposes. But also, with this kind of resolution, the lists of proteins identified from the various subcellular structures presented in section 1.3.1 of this thesis are getting increasingly organized into integrated models. This is the case for the synaptic vesicle[1] and the phagosome[2]. These models, upon which testable hypotheses for normal and pathological processes can be built, will be a prerequisite tool for the next and imminent phase of systems biology-oriented life sciences research. One of the ultimate goals resides in the integration of the information provided by the genome, the transcriptome, the proteome and the metabolome to define as completely as possible the composition of any cell or organelle in the hope to derive predictions on its behaviour[3]. Ultimately, elaboration of a

descriptive, static *in silico* whole-cell model would enable many studies including virtual fluorescence labelling, investigations of diffusion and other cell-scale processes, identification of contradictions and artefacts in the data or the production of approximate physical constants[4]. Far from being unrealistic, available bioinformatics tools already allow scaffolding of such models from which simulations can eventually be derived[4, 5]. But no matter how accurately any theoretical model is defined, it will ultimately always depend on the quality of the data used to generate it.

In this context, advances in the tools and methods employed in subcellular proteomics (reviewed in section 1.2) have been instrumental in allowing the field to reach its current status. Obviously the most important ones concern improvements in proteins and peptides separation[6, 7], peptide ionization[8, 9] and mass analysis[10, 11]. However, other procedures have not followed these examples and still remain simply inefficient. The work presented in this thesis describes the development and successful application of new and innovative methods for subcellular proteomics aiming at circumventing some of the current limitations in the field. As an alternative to detergent-based extraction of membrane proteins, which is often incomplete and selective[12], we have synthesised a new biotinylation reagent that can be used to label membrane-bound proteins and purify their proteolytic fragments in order to identify them in a gel-free approach[13]. After the proof of concept, we are now applying this procedure to the study of the very dynamic membrane of the mammalian spermatozoa. Doing so, we have already identified several interesting and unexpected protein candidates potentially involved in capacitation and fertilization.

On another front, we have circumvented problems associated with the classical density-based centrifugation by relying instead on fluorescence-assisted organelle sorting to purify corticotropes dense-core secretory granules and obtain a high quality proteome

in less than a day. The data generated highlights the importance of rapid organelle purification and of gel-free approaches in order to obtain proteomes as representative as possible of the *in vivo* state of the subcellular structure studied. More importantly, both the novel approaches can be easily adapted to other membranes or organelles. We are confident that the techniques we have developed will find broad applications and the results they have each generated will contribute significantly to the knowledge in their respective fields.

5.2 CELL SURFACE PROTEOMICS

Proteins exposed at the surface of a cell are absolutely crucial to its survival. Structure and adhesion proteins, pores, channels, receptors and others are implicated in the communication of the cell with its surroundings and in the response to changes that could affect its homeostasis. For example, there could be as many as 1000 distinct proteins playing the role of receptor at the cell surface[14]. And even if the receptorome comprises less than 5% of the human genome, it is the target of the vast majority of approved medications today[15]. Consequently, knowing which proteins are present at the surface of a cell is of the utmost importance and interest. This is particularly true for the mammalian spermatozoon whose membrane composition is very dynamic and is an indicator and determinant of its fertilization potential. Sequential exposition of cell surface proteins, either by the removal of seminal and decapacitation factors or by protein relocalization, allows progressive capacitation of the spermatozoa[16, 17], binding to the zona pellucida surrounding the oocyte[18, 19], digestion of the ZP, oocyte binding and penetration of the egg[19, 20]. Moreover, these proteins are not necessarily randomly free-

floating throughout the spermatozoa's head membrane. Some are localized to distinct regions[21], partially due to the presence of lipid rafts creating functional microenvironments[22, 23]. Clearly, the dynamics of cell surface exposed proteins is a challenge to which subcellular proteomics is well suited to respond, provided that appropriate tool for the study of membrane-bound proteins can be developed.

The avidin-biotin affinity technology has now been used for several decades[24], including for the labelling and retrieval of cell surface exposed proteins. A prototypical example of a proteomic study using this approach on human spermatozoa was described[25]. Such studies usually involve cell surface labelling, proteins extraction and separation by gel electrophoresis, followed by MS-based protein identification. In order to bypass the known limitations of extraction and gel electrophoresis as far as membrane proteins are concerned, we have designed a different approach to study the membrane proteome. Based on the design of a new biotinylation reagent, sulfo-NHS-iminobiotin, and on the proteolytic release of water soluble fragments from membrane proteins, this method was developed and a proof-of-concept study was presented in chapter 2[13]. It was further applied to the surface of freshly ejaculated spermatozoa with very encouraging preliminary results, as described in chapter 3.

Several conclusions can be drawn from the results obtained in those studies. Firstly, the biotinylation reagent we have synthesized is fully functional, water soluble and membrane impermeable. It also exhibits fully reversible, pH-dependent binding to an avidin support. As such, it is a new tool for protein labeling displaying a unique combination of physicochemical properties. Secondly, this reagent can be advantageously used as an alternative to sulfo-NHS-biotin or sulfo-NHS-SS-biotin[26] for affinity purification and complete recovery of labeled biomolecules. Indeed, the extremely strong

affinity of the biotin moiety for avidin derivatives does not allow complete subsequent elution of the bound material, even when harsh conditions and detergents are employed, leading to protein loss. This is corrected in our case by the use of an iminobiotin group allowing pH-dependent elution. On another hand, sulfo-NHS-SS-biotin allows cleavage of the label for elution but it is vulnerable to *in vivo* reducing conditions and does prevents multiple uses of the column, as opposed to our newly synthesized reagent. Also, the concept of retrieving only labeled, surface exposed, water soluble fragments of otherwise partially hydrophobic proteins circumvents all issues concerning partial and differential detergent extractions. The method is especially efficient since the technologies for peptide separation and analysis have now evolved and allow confident identification of proteins from only a few high quality peptides.

One of the important challenges in proteomics is the development of powerful bioinformatics tools and accurate databases. Until *de novo*, database-independent sequencing by mass spectrometry is available[27], the quality of protein identification will be directly related to that of the information present in the protein databases. This was certainly highlighted by our results in chapter 2. Protein identification in that case was rendered difficult because very few membrane proteins were present in the databases for our first model, *Spodoptera frugiperda* and this, despite the purification of hundreds of peptides. The quality of the purification could nevertheless be assessed by considering the contribution of peptides derived from better known soluble proteins. Another pitfall in proteomic analysis is not only the absence of information, but also the poor quality of the information contained in the various databases. Partial, redundant and erroneous entries complicate protein identification and lead to false positive results[28]. It is certainly worth mentioning that several journals in the field have recently upped their criteria for protein

identifications by mass spectrometry in order to increase the quality of the publications and the biological relevance of the presented results. Since the deciphering of the *Bos taurus* genome was well under way in 2003 and several other closely related mammalian species were under investigation, our results on bovine spermatozoa were far superior. Even if some candidates such as A-kinase anchoring proteins[29] were known and expected to be exposed at the surface, it was not the case for all proteins. For example, the presence of hexokinase[30], subunits of the proteasome[31] and others at the surface of spermatozoa have been poorly documented and this study seems to confirm those findings. Moreover, several of the proteins identified in our preliminary results have not yet been reported in the spermatozoa membrane. Once the experiments are repeated, we will be more confident about their positive identification and hence be in a position to validate their potential role in capacitation and fertilization. Also, focusing on the proteins synthesized as precursors and especially on those presenting a prohormone convertase-specific cleavage motif, we will also try to identify endogenous substrates for PC4.

The implications of our findings are also interesting on a more fundamental aspect. The use of widely accepted biological cues such as signal peptides or the presence of hydrophobic segments is fundamental to annotate proteins as secreted and/or present at the membrane. However, it is increasingly recognized that some proteins are present at the surface of cell without any apparent indication on how they reach this compartment. Among the most striking example is the finding that sex steroid nuclear receptors, long believed to act only at the DNA level, are now known to exert non-genomic (or epigenetic) actions from the cell membrane[32]. Also, many heat shock proteins, acting intracellularly in response to stress, have recently been shown to be present at the surface of normal cells[33] and even secreted from tumor cells through ill-defined non-conventional

pathways[34]. Since several of the proteins we have identified do not possess consensus secretory signals nor can they cross membranes through known transporters, questions arise on the mechanisms by which they reach the surface. The spermatozoon is a cell with a unique structure, and we can not exclude the possibility that cell-specific mechanisms of protein transport could exist to ensure proper distribution and exposition of proteins at its surface. Moreover, because of its content and the machinery involved in its formation, the acrosome is now more perceived as a specialized lysosome-derived rather than a secretory vesicle-derived structure[35]. Future proteomic results could help confirm this hypothesis or define it as being more of an hybrid of the two.

The work presented in chapters 3 and 4 could certainly be build upon in several directions. Firstly, the approach itself could be modified at the cell surface limited proteolysis step. Indeed, the parallel use of proteases other than trypsin, as long as they exhibit restricted substrate specificity, followed by independent and/or combined analysis of the lysate could generate more and different peptides to ultimately improve protein identification. Also, subcellular fractionation of the spermatozoa to obtain distinct preparations of heads and tails[36] prior to labeling could result in simpler proteomes whose analysis might provide more easily interpretable information. In this study, we have investigated the cell surface proteome of freshly ejaculated mammalian spermatozoa. However, it has been documented that membrane components of these cells in this state are very different from those exposed after capacitation and after acrosome reaction. Since both capacitation and acrosome reaction can be chemically induced *in vitro*[37], we envision that our procedure could be easily applied to all three states. Subtractive analysis of all three proteomes would then reveal proteins uniquely present in all three states, simplifying the quest for candidates involved in each step. Such an elegant proteomic

study on the dynamics of the spermatozoa membrane has not been reported to date, and our new method would certainly make such study not only feasible, but much less labor-intensive than conventional approaches. And as it is the case for both the new methods presented in this thesis, it was designed to be widely applicable. As such, it could be applied to several other models, for example, to study the alterations in cell surface proteomes throughout cell differentiation or between normal and transformed cells.

5.3 SECRETORY GRANULE PROTEOMICS

Access to improved subcellular fractionation methods is rendered almost mandatory by the fast progress in mass spectrometry-based protein identification[38]. Indeed, one of the most important objectives in sample preparation for subcellular proteomics is to have a preparation that is as highly enriched as possible for the organelle of interest. Only then can the identified proteins be assigned to that specific compartment with confidence. For several decades, gradient-based fractionation has been the tool of choice for organelle enrichment. Combined with gel-based protein separation strategies, such studies are time- and material-consuming, although still very informative. Recent proteomic studies frequently report proteins common to more than one subcellular compartment (ER and Golgi[39], secretory granule and lysosome[40]). Whereas this is expected in some cases, it is very surprising in others. Among other solutions, the development of novel subcellular fractionation and organelle purification strategies could help to confirm whether those findings are real or likely due to contamination of the preparation with other structures. On that topic, fluorescence-assisted organelle sorting (FAOS) was suggested in the 1980s[41], but technological limitations at the time prevented

efficient application of this concept. In chapter 4 of the present thesis, we revisited this approach using modern instrumentation and AtT-20 corticotropes cells expressing a mGFP-tagged protein inside their secretory granules. Since this organelle had not yet been analyzed using proteomics until a few months ago, we proposed to design, develop and optimize a new fluorescence-assisted organelle sorting protocol as well as to apply it to obtain the proteome of the endocrine secretory granule.

Concerning the method *per se,* it can be said that it compares very favorably with the established protocols, especially considering that the latter have been under continuous development for decades. Firstly, it is very rapid, allowing acquisition in a single day of a proteome from cultured cells provided that the instrumentations and facilities are available. Aside from the obvious reduction in labor and augmented throughput, it also allows, as discussed below, the identification of transiently associated proteins that would not necessarily remain associated with the organelle for the 24 to more than 48 hours required by the manipulations associated with density-based subcellular fractionation. Our purification procedure is also very efficient, allowing purification of enough secretory granules to obtain a proteome of over 100 proteins from less than $20x10^6$ cells grown in a single flask. Also, our completely gel-free procedure leads to less sample losses, including low molecular weight peptides, and higher quality MS spectra, resulting in confident protein identification with high scores. Finally, and importantly, this new approach is widely applicable to several intracellular compartments using different fluorescent molecules, ranging from fluorescent fusion proteins to dyes to conjugated antibodies. Moreover, the experimenter can easily use a combination of different fluorescent molecules for active sorting and exclusion of organelles. Also, organelles in different states, such as immature and mature secretory granules, could be

separated according to their size and the presence or absence of specific markers at their surface, such as clathrin[42]. The same approach could be used to discriminate and analyze the COPI and COPII-coated vesicles responsible for the retrograde and anterograde transport between ER and Golgi, respectively[43]. The potential of fluorescence-assisted organelle sorting is very high, provided that the separation conditions are adjusted for each subcellular compartment in order to preserve its integrity.

Using our method, we have obtained a proteome of over 150 proteins for the corticotrope secretory granule. Notable improvements over recently published secretory granule proteomes (chromaffin granules[44] and insulin-containing granules[40]) include higher protein and peptide scores, higher number of proteins identified, higher coverage of several proteins and less contamination from other subcellular compartments. The proteomic analysis of the insulin-containing granules has revealed the presence of numerous lysosomal proteins in the sample analyzed. Furthermore, their immunocytochemical validations showed at least partial colocalization of some identified lysosomal proteins with insulin leading the authors to present this result as supporting evidence of the "sorting-by-retention" model for cargo sorting to the secretory granules. Indeed, they surmized that some of the purified granules still contained material destined to be removed and rerouted to other compartments such as the lysosomes. However, despite the validations proposed and the rationale presented, it is nevertheless clear that the density centrifugation results in this case raise questions concerning the purity of the secretory granule preparation, which certainly at least, is not exclusively composed of mature dense-core secretory granule. Interestingly, the study of the chromaffin granules, on the other hand, identified very few lysosomal proteins. Even if the near-simultaneous publication of these two studies prevented them on commenting on each other, we can

hypothesize on quite a few explanations for this. First, dealing with granules isolated from an immortalized (transformed) cell line versus a normal tissue could account for some of the differences observed; an argument which, in our own study, must also be considered. Secondly, the density-based separations differed significantly, suggesting that the second study might be more reflective of the proteome of purified mature secretory granules. Indeed, the results agree more with the expected content for this organelle as well as with our own corticotropes secretory granule proteomic results employing a completely different organelle purification protocol. Finally it is not unrealistic to think that differences in the major secretory product of a cell could alter the granule's protein composition significantly.

Another striking result in our study is the identification of several granule-associated proteins involved in transport and fusion of the organelle. They are most likely transiently associated with the organelle and are probably lost during the lengthy centrifugation and electrophoresis steps of the traditional procedure. It is our conclusion that our approach can provide at least complementary, if not superior proteomic information than that obtained through classical organellar proteomic approaches. Our results tend to confirm a model for a secretory granule core largely composed of members of the granin family of proteins (chromogranin A, chromogranin B, secretogranin II, secretogranin III), the major secretory product (POMC) and its biologically active fragments and various protein processing enzymes (PC1/3, PAM). Its membrane seem to exhibit several proteins shared with the synaptic vesicle[1] (V-ATPase, exchangers, Rabs, SNAREs), consistent with their common role in intracellular cargo transport and secretion. Of great interest in our case is the association of several cytoskeletal and cytoskeletal-associated proteins (kalirin, myosin, actin, tubulin) with our granules

recovered from gentle cell lysis and FAOS. It confirms that secretory granules are not free floating in the cell, but rather strongly tethered along microtubules and within actin filament meshes as recently evidenced using photonic force microscopy[45]. Our results might therefore be more representative of the *in vivo* state of the organelle. The use of cytoskeletal disruptors and freeze-thaw cycles for cell lysis has proven, in our case, less disruptive leading to diminished cross-contamination than standard Potter homogenization. We suggest that use of the latter method should be revisited and might not be optimal for the isolation of a pure preparation of intact organelles for proteomic analysis.

Finally, several new proteins were identified whose role in secretory granule's biology is not clearly defined. 14-3-3, a scaffolding protein and cyclophilin A, a prolyl cis-trans isomerase could help in cargo structure and polymerization. Other proteins, such as M2-pyruvate kinase[46], chaperones and heat shock proteins[47] are present in the plasma of patients with endocrine tumors, although the mechanism by which they are secreted remains undefined. Since AtT-20 cells are immortalized, it might represent a model to study how some of these unexpected proteins, some of them without a signal peptide, find their way into the regulated pathway of secretion. Further validation, through immunocytochemical and stimulated secretion studies for example, will be necessary before the role of the newly identified proteins can be further studied.

Comparison of our results with that of other proteomic studies on the endocrine secretory granules suggests that they could contain both common and distinct proteins depending on cell type and cargo. Granule-to-granule heterogeneity has already been observed within a single cell[48]. Whether molecular mechanisms exist to generate or recognize these differences and use them to control regulated secretion more precisely,

for example, by piecemeal degranulation[49], is still unknown. Elucidation of proteomes of other types of endocrine granules and secretory vesicles is certainly a valid approach to investigate this topic.

Once again, the method presented in chapter 4 could easily be adapted to the purification and proteomic analysis of other organelles, making it usable by any group having access to a flow cytometer. In our case, one of the most interesting follow-up studies would certainly be the application of our strategy to other types of endocrine secretory granules. As mentioned in the introduction, there are several commercially available cell lines for each cell type of the anterior pituitary[50]. Transfection of those cell lines with PHM-mGFP, purification of their granules by FAOS and analysis of their proteomes could rapidly provide biologically relevant information. Analysis of the proteins common to all granules could provide insights on the mechanisms of granule biogenesis, maturation and secretion, while candidates unique to each cell type could yield information on the sorting, processing and action of their individual major secretory product.

Moreover, both methods presented in this thesis are not only widely applicable on their own, but could easily be used in combination. The membrane-bound and luminal portions of the organelles' proteomes could then be distinguished.

The field of proteomics has progressed at a tremendous pace in the last few years. From the simple description of the protein content of whole cells and organisms, it now focuses on the precise and dynamic description of subcellular compartments and the bioinformatics integration of all sources of information in the elaboration of whole-organelle models. The new methods presented in this thesis are certainly invaluable tools to tackle two current limitations in subcellular proteomics, namely the non-discriminatory

identification of membrane proteins and the fast and efficient purification of intracellular structures. Subcellular proteomics now provides more than exhaustive protein lists, it has become a most efficient strategy to identify disease-causing targets and describe and monitor large-scale changes at the protein level in any biological system. The improvements in available technologies, both applied and computational, together with the emergence of the field of systems biology allow one to predict that relatively accurate whole-cell *in-silico* simulations could be available in a near future. Such tools would be groundbreaking as far as the development of therapies is concerned.

5.4 CLAIMS TO ORIGINALITY

Chapter 2

1. Chemical synthesis of a novel biotinylation reagent, sulfo-NHS-iminobiotin possessing a unique combination of physicochemical properties. They include water-solubility, membrane-impermeability, covalent modification of primary amines and fully reversible, pH-dependent binding to avidin derivatives under mild conditions.
2. Successful application of sulfo-NHS-iminobiotin in the development of a new, detergent-, extraction- and gel-free protocol for the acquisition and analysis of surface-exposed, membrane associated proteins. The protocol presents a distinct combination of cell surface labeling, cell surface limited proteolytic digestion and membrane protein fragments recovery and purification in near-native conditions.
3. Identification of several *Spodoptera frugiperda* peptides from cell surface proteins not previously reported in the databases.
4. As mentioned previously, these claims would have been substantiated through a patent to DJ Gauthier and C Lazure if it had not been judged unsuitable due to a simple time technicality.

Chapter 3

1. Partial biochemical and proteomic characterization of some of the PC4-like immunoreactive forms.
2. Application of our new non-discriminatory investigative approach to the study of cell surface proteins of freshly ejaculated bovine spermatozoa.
3. Proteomic identification of hundreds of mammalian spermatozoa surface proteins, many of which never previously reported in this model.

Chapter 4

1. Design of a new, centrifugation- and gel-free method for the purification and proteomic analysis of corticotropes secretory granules readily applicable to other organelles.
2. Optimization of the fluorescence-assisted organelle sorting protocol to the corticotropes secretory granules and to the further analytical steps in proteomic analysis.
3. FAOS of the corticotropes secretory granules and proteomic analysis of their protein composition, thereby establishing the most complete secretory granule proteome to date.
4. Identification of several new secretory granule proteins, including transiently associated proteins not previously identified by classical protocols.

5.5 BIBLIOGRAPHY

1. Takamori, S. et al. Molecular anatomy of a trafficking organelle. *Cell* **127**, 831-846 (2006).

2. Stuart, L.M. et al. A systems biology analysis of the Drosophila phagosome. *Nature* **445**, 95-101 (2007).

3. Dettmer, K., Aronov, P.A. & Hammock, B.D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51-78 (2007).

4. Betts, M.J. & Russell, R.B. The hard cell: From proteomics to a whole cell model. *FEBS Lett.* (2007).

5. Tomita, M. Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol.* **19**, 205-210 (2001).

6. Gorg, A., Weiss, W. & Dunn, M.J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685 (2004).

7. Jessani, N. et al. A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods* **2**, 691-697 (2005).

8. Koichi Tanaka, H.W.Y.I.S.A.Y.Y.T.Y.T.M. Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151-153 (1988).

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

10. Wysocki, V.H., Resing, K.A., Zhang, Q. & Cheng, G. Mass spectrometry of peptides and proteins. *Methods* **35**, 211-222 (2005).

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

12. Santoni, V., Molloy, M. & Rabilloud, T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070 (2000).

13. Gauthier, D.J., Gibbs, B.F., Rabah, N. & Lazure, C. Utilization of a new biotinylation reagent in the development of a nondiscriminatory investigative approach for the study of cell surface proteins. *Proteomics* **4**, 3783-3790 (2004).

14. Ben-Shlomo, I., Yu Hsu, S., Rauch, R., Kowalski, H.W. & Hsueh, A.J. Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci. STKE* **2003**, RE9 (2003).

15. Strachan, R.T., Ferrara, G. & Roth, B.L. Screening the receptorome: an efficient approach for drug discovery and target validation. *Drug Discov. Today* **11**, 708-716 (2006).

16. Fraser, L.R., Adeoya-Osiguwa, S.A., Baxendale, R.W. & Gibbons, R. Regulation of mammalian sperm capacitation by endogenous molecules. *Front. Biosci.* **11**, 1636-1645 (2006).

17. Nixon, B. et al. The identification of mouse sperm-surface-associated proteins and characterization of their ability to act as decapacitation factors. *Biol. Reprod.* **74**, 275-287 (2006).

18. Talbot, P., Shur, B.D. & Myles, D.G. Cell adhesion and fertilization: steps in oocyte transport, sperm-zona pellucida interactions, and sperm-egg fusion. *Biol. Reprod.* **68**, 1-9 (2003).

19. Snell, W.J. & White, J.M. The molecules of mammalian fertilization. *Cell* **85**, 629-637 (1996).

20. Rubinstein, E., Ziyyat, A., Wolf, J.P., Le Naour, F. & Boucheix, C. The molecular players of sperm-egg fusion in mammals. *Semin. Cell Dev. Biol.* **17**, 254-263 (2006).

21. Flesch, F.M. & Gadella, B.M. Dynamics of the mammalian sperm plasma membrane in the process of fertilization. *Biochim. Biophys. Acta* **1469**, 197-235 (2000).

22. Cross, N.L. Reorganization of lipid rafts during capacitation of human sperm. *Biol. Reprod.* **71**, 1367-1373 (2004).

23. Thaler, C.D., Thomas, M. & Ramalie, J.R. Reorganization of mouse sperm lipid rafts by capacitation. *Mol. Reprod. Dev.* **73**, 1541-1549 (2006).

24. Wilchek, M. & Bayer, E.A. Applications of avidin-biotin technology: literature survey. *Methods Enzymol.* **184**, 14-45 (1990).

25. Shetty, J. et al. Differential extraction and enrichment of human sperm surface proteins in a proteome: identification of immunocontraceptive candidates. *Electrophoresis* **22**, 3053-3066 (2001).

26. Shimkus, M., Levy, J. & Herman, T. A chemically cleavable biotinylated nucleotide: usefulness in the recovery of protein-DNA complexes from avidin affinity columns. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2593-2597 (1985).

27. Palagi, P.M., Hernandez, P., Walther, D. & Appel, R.D. Proteome informatics I: bioinformatics tools for processing experimental data. *Proteomics* **6**, 5435-5444 (2006).

28. Biron, D.G. et al. The pitfalls of proteomics experiments without the correct use of bioinformatics tools. *Proteomics* **6**, 5577-5596 (2006).

29. Carr, D.W. & Newell, A.E. The role of A-kinase anchoring proteins (AKaps) in regulating sperm function. *Soc. Reprod. Fertil. Suppl.* **63**, 135-141 (2007).

30. Visconti, P.E. et al. Properties and localization of a tyrosine phosphorylated form of hexokinase in mouse sperm. *Mol. Reprod. Dev.* **43**, 82-93 (1996).

31. Bohring, C. & Krause, W. The role of antisperm antibodies during fertilization and for immunological infertility. *Chem. Immunol. Allergy* **88**, 15-26 (2005).

32. Peluso, J.J. Non-genomic actions of progesterone in the normal and neoplastic mammalian ovary. *Semin. Reprod. Med.* **25**, 198-207 (2007).

33. Pfister, G. et al. Detection of HSP60 on the membrane surface of stressed human endothelial cells by atomic force and confocal microscopy. *J. Cell Sci.* **118**, 1587-1594 (2005).

34. Mambula, S.S. & Calderwood, S.K. Heat shock protein 70 is secreted from tumor cells by a nonclassical pathway involving lysosomal endosomes. *J. Immunol.* **177**, 7849-7857 (2006).

35. Moreno, R.D. & Alvarado, C.P. The mammalian acrosome as a secretory lysosome: new and old evidence. *Mol. Reprod. Dev.* **73**, 1430-1434 (2006).

36. Stein, K.K., Go, J.C., Lane, W.S., Primakoff, P. & Myles, D.G. Proteomic analysis of sperm regions that mediate sperm-egg interactions. *Proteomics* **6**, 3533-3543 (2006).

37. Wu, J.T., Chiang, K.C. & Cheng, F.P. Expression of progesterone receptor(s) during capacitation and incidence of acrosome reaction induced by progesterone and zona proteins in boar spermatozoa. *Anim. Reprod. Sci.* **93**, 34-45 (2006).

38. Yates, J.R., 3rd, Gilchrist, A., Howell, K.E. & Bergeron, J.J. Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell. Biol.* **6**, 702-714 (2005).

39. Gilchrist, A. et al. Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265-1281 (2006).

40. Brunner, Y. et al. Proteomic analysis of insulin secretory granules. *Mol. Cell. Proteomics* (2007).

41. Murphy, R.F. Analysis and isolation of endocytic vesicles by flow cytometry and sorting: demonstration of three kinetically distinct compartments involved in fluid-phase endocytosis. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 8523-8526 (1985).

42. Kakhlon, O., Sakya, P., Larijani, B., Watson, R. & Tooze, S.A. GGA function is required for maturation of neuroendocrine secretory granules. *EMBO J.* **25**, 1590-1602 (2006).

43. Lee, M.C., Miller, E.A., Goldberg, J., Orci, L. & Schekman, R. Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **20**, 87-123 (2004).

44. Wegrzyn, J., Lee, J., Neveu, J.M., Lane, W.S. & Hook, V. Proteomics of Neuroendocrine Secretory Vesicles Reveal Distinct Functional Systems for Biosynthesis and Exocytosis of Peptide Hormones and Neurotransmitters. *J. Proteome Res.* **6**, 1652-1665 (2007).

45. Abu-Hamdah, R., Cho, W.J., Horber, J.K. & Jena, B.P. Secretory vesicles in live cells are not free-floating but tethered to filamentous structures: a study using photonic force microscopy. *Ultramicroscopy* **106**, 670-673 (2006).

46. Ahmed, A.S. et al. M2-PK as a novel marker in ovarian cancer. A prospective cohort study. *Eur. J. Gynaecol. Oncol.* **28**, 83-88 (2007).

47. Gastpar, R. et al. Heat shock protein 70 surface-positive tumor exosomes stimulate migratory and cytolytic activity of natural killer cells. *Cancer Res.* **65**, 5238-5247 (2005).

48. Sobota, J.A., Ferraro, F., Back, N., Eipper, B.A. & Mains, R.E. Not all secretory granules are created equal: Partitioning of soluble content proteins. *Mol. Biol. Cell* **17**, 5038-5052 (2006).

49. Crivellato, E., Nico, B., Bertelli, E., Nussdorfer, G.G. & Ribatti, D. Dense-core granules in neuroendocrine cells and neurons release their secretory constituents by piecemeal degranulation (review). *Int. J. Mol. Med.* **18**, 1037-1046 (2006).

50. Ooi, G.T., Tawadros, N. & Escalona, R.M. Pituitary cell lines and their endocrine applications. *Mol. Cell. Endocrinol.* **228**, 1-21 (2004).