2010

Characterization of cellular stress systems using biological mass spectrometry

Anna Kozarova
University of Windsor

Characterization of cellular stress systems using biological mass spectrometry

by
Anna Kozarova

A Dissertation
Submitted to the Faculty of Graduate Studies
through Chemistry and Biochemistry
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2010

Your file Votre référence
ISBN: 978-0-494-80248-9
Our file Notre référence
ISBN: 978-0-494-80248-9

DECLARATION OF CO-AUTHORSHIP / PREVIOUS PUBLICATION

I. Co-Authorship Declaration

I hereby declare that this thesis incorporates material that is the result of joint research, as follows:

This thesis incorporates the outcome of joint research undertaken in collaboration with Michelle A. Sharon under the supervision of Dr. Alden H. Warner, Department of Biological Sciences. The collaboration is covered in Chapter 2 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the authors; Michelle purified a 21 kDa protein from *Artemia franciscana*, PCR amplified cDNA from an embryonic library corresponding to its gene and performed the initial characterization of this protein. My contribution to this manuscript was the *de novo* identification of this protein using mass spectrometry (MS). Since the genome of this organism has not yet been sequenced, this included *in silico* analysis for the identification. Based on the peptide sequence corresponding to a conserved motif, the 21 kDa protein was identified as a member of group 1 Late embryogenesis abundant (LEA) family. Once Michelle completed the gene amplification, I was able to match the peptide sequences generated by MS and thus further confirm the identity of the 21 kDa protein.

This thesis also incorporates the outcome of joint research undertaken in collaboration with Inga Sliskovic under the supervision of Dr. Bulent Mutus, Department of Chemistry and Biochemistry. The collaboration is covered in Chapter 3 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the authors, and the contribution of co-authors was primarily through the provision of recombinant purified protein disulfide isomerase, PDI.

Inga prepared the PDI protein sample in a reduced and in an auto-oxidized form. Dr. Eric Simon and Dr. Philip Andrews at the Michigan Proteome Consortium, University of Michigan, provided knowledge and instrumentation for the peptide separation and MALDI TOF/TOF analysis. My contribution to this manuscript was application of the isotope-coded affinity tag (ICAT) technology and identification of the cysteine redox state on PDI using mass spectrometry (MS). I performed all experimental steps involved in the ICAT labeling protocols, followed with acquisition and interpretation of the MS spectra.

This thesis also incorporates the outcome of joint research undertaken in collaboration with Dr. John W. Hudson, Department of Biological Sciences. The collaboration is covered in Chapter 4 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the authors, and the contribution of co-authors was through the provision of experimental help, analysis and data interpretation with flow cytometry based approach. My contribution to this manuscript in preparation was identification of phosphorylation sites on Flag-hYVH1 expressed and immunoprecipitated from mammalian cells; generation of phosphomutants, localization experiments, RNAi experiments and characterization of the phenotype, as well as sample processing for flow cytometry experiments.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II. Declaration of Previous Publication

This thesis includes 2 original papers that have been previously published/submitted for publication in peer reviewed journals, and 1 manuscript in preparation as follows:

Thesis Chapter	Publication title/full citation	Publication status*
Chapter 2	Sharon, M.A., Kozarova, A., Clegg, J.S., Vacratsis, P.O., and Warner, A.H. (2009). Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp Artemia franciscana. Biochem Cell Biol *87*, 415-430	Published
Chapter 3	Kozarova, A., Sliskovic, I., Mutus, B., Simon, E.S., Andrews, P.C., and Vacratsis, P.O. (2007). Identification of redox sensitive thiols of protein disulfide isomerase using isotope coded affinity technology and mass spectrometry. J Am Soc Mass Spectrom *18*, 260-269	Published
Chapter 4	Kozarova, A., Hudson, J.W., Vacratsis, P.O.	To be submitted

I certify that I have obtained written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques,

quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

In recent years mass spectrometry has become an invaluable tool to address an array of biological questions. The versatility of this experimental approach does not only allow assignment of protein identity and identification of sequence specific modifications, but with the help of particular derivatization techniques facilitates the determination of peptide quantity. Each of these approaches were applied to the following biological projects:

The 21 kDa heat stable protein purified from the encysted embryo of *Artemia franciscana* was characterized by time-of-flight mass spectrometry. *De novo* sequencing of peptides identified this protein as a group 1 Late embryogenesis abundant (LEA) protein. The amino acid sequence assignment to these peptides allowed amplification of the entire gene sequence from an embryonic cDNA library. This was deposited into the NCBI database (EF656614). The expression of group 1 LEA protein is consistent with and supports a role in desiccation tolerance. In addition, this is a first report describing identification of a group 1 LEA protein in an animal species.

A MS-based quantitative analysis was performed in order to analyze relative changes in the dynamic thiol and disulfide states of the redox sensitive protein disulfide isomerase, PDI. PDI cysteine residues were derivatized with an isotope-coded affinity tag (ICAT), thus allowing a direct comparison between the reduced and auto-oxidized forms. Quantitation was based on relative ratios between light and heavy isotopically labeled cysteine containing peptides. The application of the ICAT-labeling approach to PDI related studies, allowed direct assignment of individual cysteine residues and their oxidation status, compared to the previously employed techniques, that only provided

information regarding the average number of modified cysteine residues within PDI, not their exact identity.

A combination of a phosphopeptide enrichment step and a MS-based approach was utilized to identify three phosphorylation sites on hYVH1, an atypical dual specificity phosphatase that functions as a cell survival factor. With the help of phosphomimetic and non-phosphorylable mutants, we were able to decipher their effect on localization and progression through the cell cycle.

Collectively, these studies manifest the power of MS-generated data to influence and guide many different fields ranging from molecular embryology to biochemistry.

DEDICATION

To my family

ACKNOWLEDGMENT

First, and most importantly, I would like to thank my supervisor, Dr. Otis Vacratsis, for giving me the opportunity to work with him and for sharing his personal "tricks of the trade" related to mass spectrometry. I would also like to thank him for the hours long scientific discussions, help and patience, and for the words of encouragement whenever experiments went sour. Overall, thank you Otis for providing a great environment for discovery in the complicated world of science.

I would also like to thank the members of my committee; Drs. Andrew Hubberstey, Lana Lee and Siyaram Pandey for their valuable discussions. I would like to thank my collaborators, Drs. Bulent Mutus and Alden Warner, for the ability to work together on their projects while acquiring extraordinary technical experience. I would also like to thank Dr. John Hudson, for all his help during my tenure as a graduate student.

Additionally, I would like to express words of appreciation to all members of the Chemistry and Biochemistry Department who made their equipment available to me. A special thank you is in place for Dr. Michael Crawford from the Department of Biological Sciences, for letting me use his microscope. Additionally, I thank Drs. Eric Simon and Philip Andrews at the Michigan Proteome Consortium, University of Michigan, for providing knowledge and access to high end mass spec instruments. Last, but not least, I would like to thank all members of the Vacratsis lab, present and past, for a good working relationship and friendship.

The members of my family deserve an award for their support and understanding during my time in the graduate school and writing of this thesis. I am very grateful to everyone, who supported me and understood when research came first.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVATIONS

ACN	acetonitrile
ATM	ataxia telangiectasia mutated
ATR	ATM- and Rad3 related
BSA	bovine serum albumin
CAF	chemically assisted fragmentation
Cal A	calyculin A
CDK	Cyclin-dependent kinase
CG	cytosine-guanine
CID	collision-induced dissociation
CIP	calf intestinal alkaline phosphatase
CsA	cyclosporin A
DiFMUP	6,8-diFluoro-4-methylumbelliferyl phosphate
DNA-PK	DNA-dependent protein kinase
DTT	dithiotreitol
DUSP	dual specificity phosphatase
EGTA	bis(aminoethyl)glycolether-N,N,N',N'-tetraacetic acid
ERK	Extracellular signal-regulating kinase
ESI	electrospray ionization
HPLC	high pressure liquid chromatography
HRP	horseradish peroxidase
Hsp70	heat shock protein 70
ICAT	isotope-coded affinity tags
IgG	immunoglobulin G
IMAC	immobilized metal affinity chromatography
iTRAQ	isobaric tags for relative and absolute quantification
JNK	c-Jun NH_2 terminal kinase
LEA	Late embryogenesis abundant
m/z	mass-to-charge ratio
MALDI	matrix assisted laser desorption/ionization
MALDI-TOF	matrix assisted laser desorption/ionization time-of-flight
MAPK	Mitogen activated protein kinase
MOAC	metal oxide affinity chromatography
MS	Mass Spectrometry
MudPIT	multidimensional chromatography
NEM	N-ethylmaleimide
OA	okadaic acid
PAI	protein abundance index
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PSD	post-source-decay
PTP	protein tyrosine phosphatase
RNAi	RNA interference
SCX	strong cation exchange chromatography
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis

siRNA	small interfering RNA
TAP	tandem affinity purification
TBST	tris-buffered saline containing 0.1% Tween-20
TOF	time-of-flight
TP	threonine proline

CHAPTER 1

Literature Review

Recent advances in mass spectrometry have brought this technique to the forefront of discovery based system biology research. Since a major theme of this dissertation is the application of different mass spectrometry based approaches to answer significant biological questions, an overview of several different mass spectrometry based approaches will be discussed, including methods employed during this dissertation.

ABC's of mass spectrometry

Mass spectrometry (MS) in recent years has become a fundamental experimental approach in biochemistry. Mass spectrometers by definition measure mass-to-charge ratio (m/z) of ionized analytes (Figure 1) [1]. A mass spectrometer is characterized by three major components, an ion source, a mass analyzer and an ion detector (Figure 2A).

Originally, MS was restricted to measurement and analysis of small and thermostable molecules, because large organic molecules (such as peptides and proteins) would undergo extensive fragmentation and decomposition upon vaporization into the gas phase [2]. This was overcome by the development of "soft" ionization methods (where ions are created with low internal energy and thus undergo limited fragmentation), such as electrospray ionization (ESI) [2] and matrix assisted laser desorption/ionization (MALDI) [3]. These two ionization methods are fundamentally different. In the case of ESI, a liquid sample is injected through a needle and the surface of the emerging liquid is charged by an electric field, thus forcing it into a fine spray of charged droplets [2], thereby producing a continuous beam of ions (Figure 2B) [4]. In the case of MALDI, the gas phase ions are produced from a dry co-crystallized liquid sample with an organic matrix [3]. The energy from the laser is deposited into the matrix causing rapid thermal heating and thus generation of ions of the analyte molecules in a gas phase [3], thereby

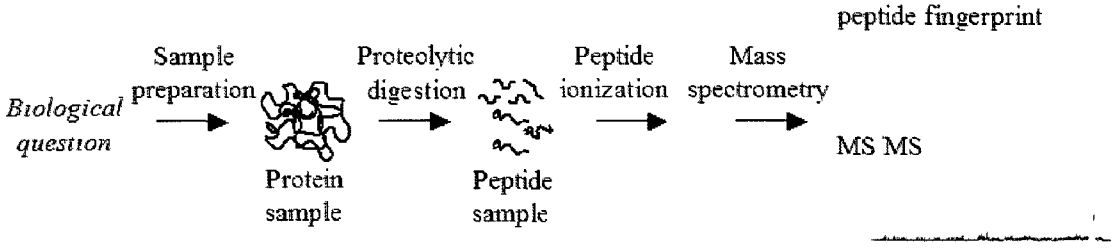


Figure 1. A simplified version of a Mass spectrometry (MS) experiment
A protein sample is prepared by an appropriate purification protocol dependent on both, the type of biological question being asked, and the sample source (*e.g.* cultured cells, bacterially purified protein or tissue samples). The sample is then subjected to enzymatic proteolysis by common proteases such as trypsin, GluC, Lys-C and Asp-N. Generated peptides are then ionized by electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI) for subsequent analysis by the mass spectrometer. Ionized peptides are detected in MS mode as a peptide fingerprint and the MS/MS mode provides information regarding the amino acid sequence of a particular peptide.

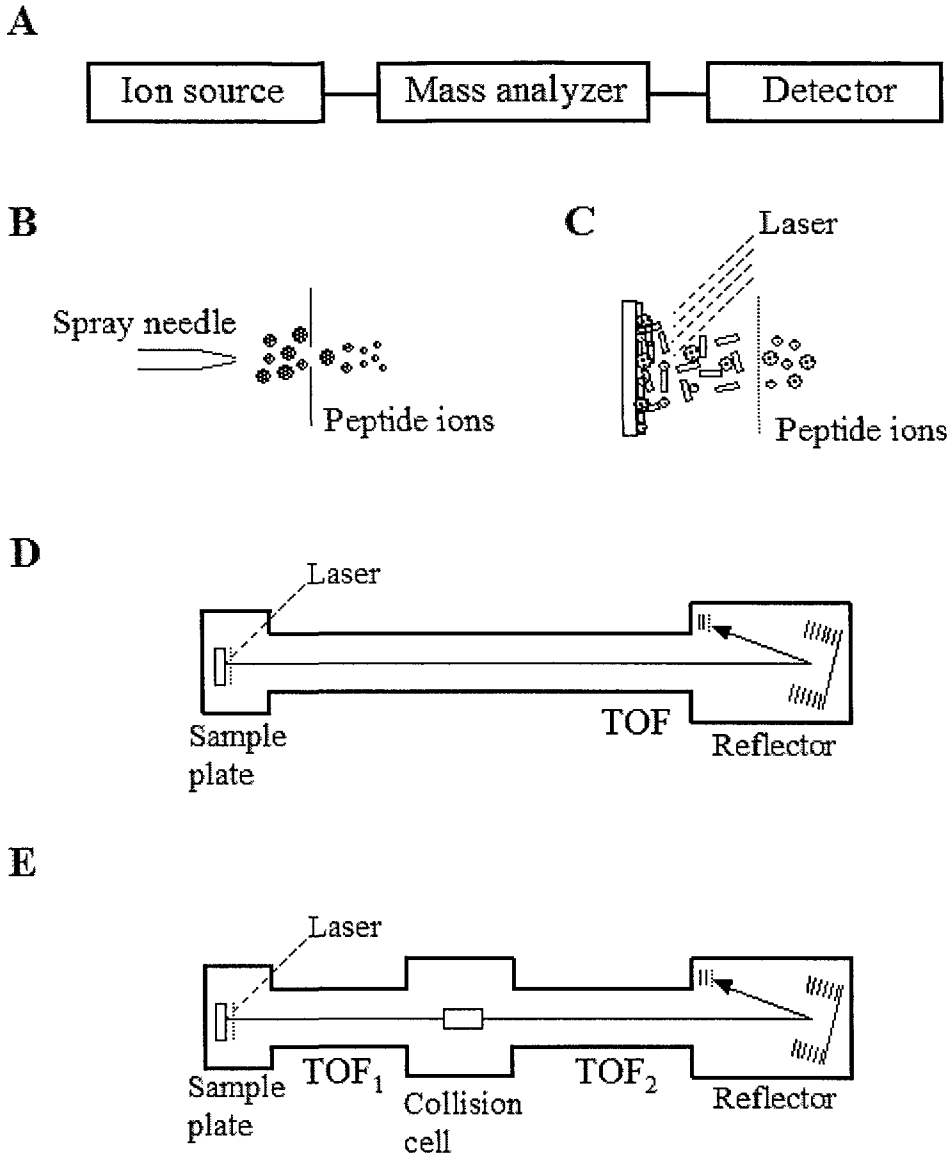


Figure 2. Basics of peptide ionization and time-of-flight instruments
(**A**) The three major components of a mass spectrometer. (**B**) A schematic depiction of the ionization process in a typical electrospray instrument, ESI. (**C**) A schematic depiction of the ionization process by matrix-assisted laser desorption/ionization, MALDI. (**D**) A simplified schematic of a MALDI time-of-flight (TOF) instrument. The ions are first accelerated to high kinetic energy allowing separation based on their different velocities along the flight tube. The detector amplifies and records arriving ions. (**E**) A simplified schematic of a MALDI TOF/TOF instrument. A collision cell has been incorporated between the two TOF sections.

producing packets of ions (Figure 2C) [4]. Another common distinction for peptide analysis is that mainly singly charged ions are generated upon MALDI, whereas ESI generates multiply charged species [5].

There are currently four basic types of mass analyzers for the study of peptides and proteins; time-of-flight (TOF), ion trap, quadrupole and Fourier transform ion-cyclotron (FTICR) [1, 6, 7]. The separation of ions in these mass analyzers is based on flight time (TOF instruments), stability (quadrupoles) and resonance frequency (ion traps, orbitraps, FTICR instruments) [5]. Each analyzer has a characteristic behaviour with regards to resolving power, sensitivity and dynamic range. In a quest to make the analyzers better and more versatile, a new generation of 'hybrid' instruments have been introduced. These instruments containing a parallel configuration of more than one analyzer, can be either identical, as in case of TOF-TOF, or different as for examples in quadrupole TOF, linear ion trap-orbitrap, *etc.* [1, 5, 6, 8]. The detector, the last but not the least important part of the mass spectrometer, monitors voltage pulses and converts these into a signal. This signal represents an ion intensity for a particular m/z and is displayed as mass spectra [7]. As the work described in this dissertation was performed on MALDI-TOF and MALDI-TOF/TOF mass spectrometers, only these instruments will be described in further details.

In the TOF instrument, the m/z is determined from flight times of peptides through a tube of specified length under vacuum. This measures the mass of intact peptides, which have been generated from intact proteins by enzymatic proteolysis prior to analysis. The major difference between the MALDI-TOF (Figure 2D) and MALDI-TOF/TOF instrument (Figure 2E) is in how the information regarding the amino acid sequence (MS/MS) of a specific peptide is generated. In a MALDI-TOF instrument, the

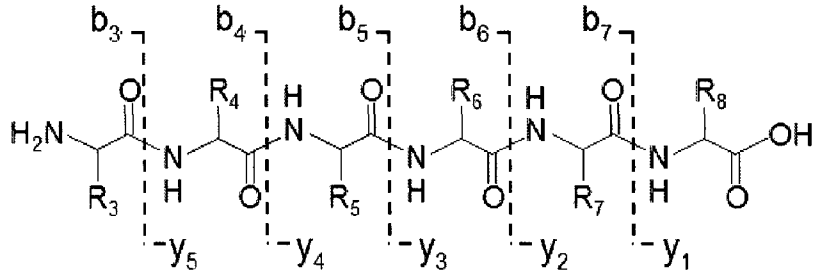
primary sequence analysis can be obtained by a process called post-source-decay (PSD), which is based on metastable fragmentation. PSD can be observed when the reflecting electrical field (reflectron) is placed at the end of the flight tube [9]. Upon selection of the parental ion $[MH]^+$ for analysis, a MALDI-PSD spectrum is obtained as a stitch of manually acquired spectra corresponding to different portions of the peptide under investigation (from the parental ion to the immonium ions) in a form of windows. Each window (correlates to a part of the spectra) corresponds to gradual changes in reflectron voltages [9]. A more common method to induce fragmentation for MS/MS analysis is collision-induced dissociation (CID), which is based on multiple collisions between a peptide under study with inert gas atoms [10]. The MALDI TOF-TOF instrument contains a collision cell between two TOF sections (Figure 2E). Upon selection of the parental ion $[MH]^+$ the entire MS/MS spectrum is automatically acquired.

The MS/MS spectrum of a peptide precursor ion allows the assignment of an amino acid sequence. As the amide bond is the lowest energy bond in peptides, this is where most of the fragmentations will occur, thus generating y-ions (C-terminal fragment ions) and b-ions (N-terminal fragment ions) (Figure 3A) [7]. Additionally, internal fragments due to double cleavage of peptide bond may also be observed [7].

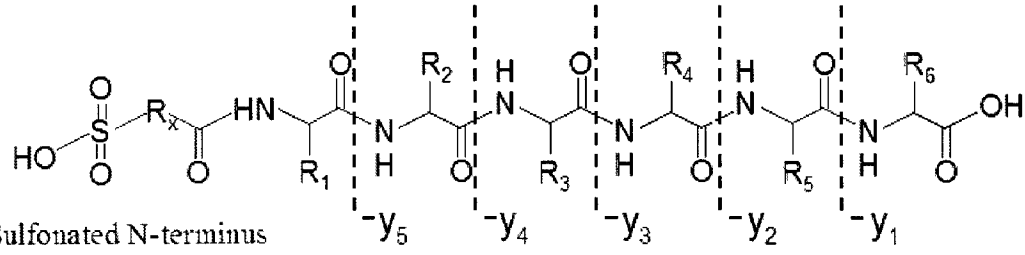
Mass spectrometry and biological questions

Biological mass spectrometry mainly focuses on the complex world of proteins. This diversity is not only due to the fact that a particular gene can give rise to multiple distinct proteins as a result of alternative splicing or that some proteins can differ in a few amino acids due to sequence polymorphisms, but it is also a reflection of the vast array of post-translational modifications that cause further complexity; such as phosphorylation,

A



B



Sulfonated N-terminus
(+136 amu)

Figure 3. The ABC's of peptide sequencing
(**A**) Within the MS-instrument, the bond breakage mainly occurs at the amide bond, the lowest energy bond. If the charge is retained by the amino-terminal fragment, b-ions are formed and detected; if the charge is retained by the carboxy-terminal fragment, y-ions are formed and detected. (**B**) Derivatization of a peptide by chemically assisted fragmentation (CAF). CAF reagent introduces a negatively charged group to the amino-terminus of tryptic peptides, thus only y-ions will be detected in positive ion mode.

glycosylation and ubiquitination just to mention a few. The changes in post-translational modifications can be either analyzed conceptually for the entire proteome (proteomics) or for a single protein. This will be described later.

With the advancement of the instrumentation and analysis software, one could now answer questions like what proteins are present within the yeast 80S ribosome [11], what dynamic changes in the complex composition does the pre-60S ribosome undergo [12] and what proteins are residents in which mammalian organelle [13]? These are a just few examples of the many proteomics studies reported in the last ten years. One limitation of this technology is the under-representation of low abundance proteins. Furthermore, as MS-based identifications rely on the detection of peptides, additional factors must be considered including that some peptides might be too small or too large to be detected, not all peptides are ionized efficiently and some might saturate the detector [14]. In order to lessen experimental bias due to the presence and ionization of high abundance peptides, a selected reaction monitoring (SRM)-based MS has been recently implemented [15-17]. In this experimental approach, the parameters for MS/MS data acquisition are pre-set, so that the instrument will only record three to five unique proteotypic peptides for each protein, if these correspond to calculated y-ion series under two main charge states. Furthermore, in a multiple reaction monitoring (MRM)-based MS, multiple peptides can be targeted in a single experiment by repeatedly cycling through a list of SRM-ion pairs associated with a set of specific retention times [18].

One of the major forces enabling proteomic analyses was the completion of entire genome sequences for numerous organisms. These allow comparison between acquired MS and MS/MS spectra with theoretically calculated fragment sizes from the *in-silico* digest of a protein sequence in the databases. This is feasible because proteases

commonly used to digest proteins into peptide are sequence specific, for example trypsin only cleaves C-terminal to lysine and arginine residues.

However, to date, not all of the organisms have their genomes sequenced and not all of the proteins sequences have been deposited into databases. As the annotations of MS/MS spectra containing y- and b-ions without prior sequence information are challenging, simplifications of the spectra can be achieved by chemical derivatization (so called chemically assisted fragmentation, CAF) [19, 20]. This method allowing *de novo* sequencing is based on attachment of sulfonic acid to the N-terminus of each peptide (+136 Da), which promotes an efficient charge-site initiated cleavage of peptide amide bonds, and thus enables selective detection of a single series of fragment ions, the C-terminal y-ions (Figure 3B) [19, 20]. A step to protect lysine side chains from sulfonation includes conversion to homoarginine (+42 Da) [20].

One of the organisms for which the genome has not been sequenced yet is the brine shrimp *Artemia franciscana*. In collaboration with Dr. A. Warner, Department of Biological Sciences, University of Windsor we employed a MS-based approach to identify a novel protein from this species. A heat-stable protein was purified from encysted *A. franciscana* embryos and identified from peptide fragments by CAF derivatization combined with *de novo* sequencing. This study will be discussed in detail in Chapter 2 of this thesis.

Biological mass spectrometry and sample complexity

The complexity of peptide samples can be decreased by an array of separation techniques. Peptides generated from proteins digested in solution (the "bottom-up" approach, [1] can be fractionated by multidimensional chromatography called MudPIT,

where the first separation is based on charge (strong cation exchanger) and the second on hydrophobicity (reverse-phase) prior to direct injection into the electrospray based MS-instruments (Figure 4A) [11, 21].

Alternatively, instead of decreasing complexity of peptide samples by chromatographic separation, proteins can be first separated on one-dimensional (1D) or two-dimensional (2D) SDS-PAGE prior to enzymatic digestion into peptides. This allows identification by either MALDI or ESI-MS (Figure 4B).

An elegant protocol based on the use of tandem affinity purification (TAP) tags allows co-purification of interacting partners at substoichiometric levels [22]. Originally this method included an SDS-PAGE separation of the protein components of the immunoprecipitated complex and the collection of unique protein bands allowing separate proteolytic digests to be performed. Overall, this decreased the complexity of analyzed peptide samples [22]. The classical route for identifying interacting partners involves over-expression systems, that usually employ genetically engineered fusion proteins containing one tag, for example Flag-tag, Myc-tag, *etc.* These tags allow affinity purification and/or detection, when a suitable antibody directed against endogenous protein is not available, or not suitable for these applications. However, the ability to detect fusion proteins at low concentrations in a complex mixture in these one-tag over-expression systems is limited. This problem was circumvented by the introduction of two different tags (the TAP-tag system) separated by a protease recognition sequence (tev) in frame with the protein for which the goal is to identify interacting partners (Figure 4C). After expression of this fusion protein, the first purification step is based on the affinity between IgG –binding domain (the first tag) and the IgG beads. Upon removal of non-specific proteins by washing steps, the fusion protein is released from the beads upon

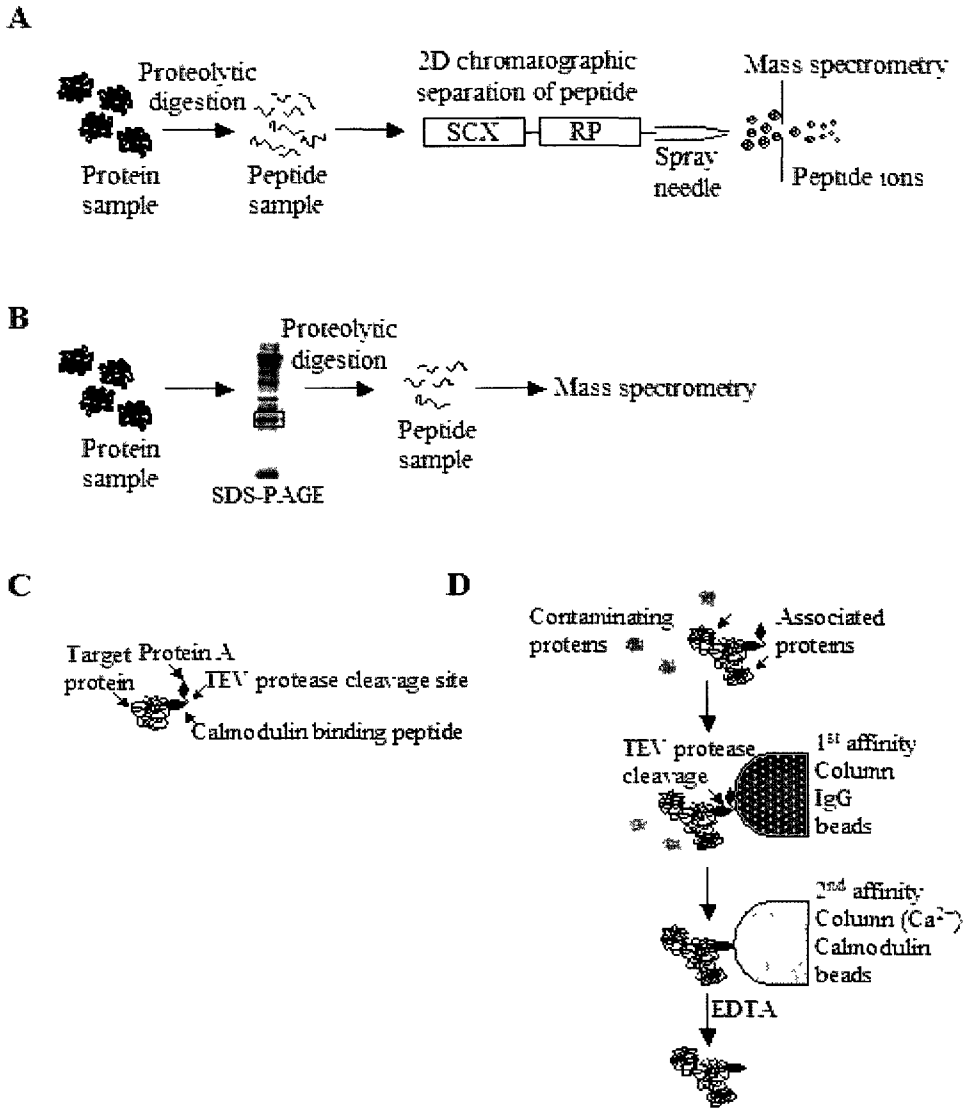


Figure 4. Alternative approaches to decrease sample complexity

(**A**) Peptide fractionation by the MudPIT, multidimensional chromatography method [11, 21]. After enzymatic proteolysis in solution, acidified peptides are first separated on a strong cation exchange column (SCX). Then as a discrete fraction of peptides are displaced from this column, peptides are further resolved by a reverse phase (RP) chromatography column followed by direct injection into the MS-instrument. (**B**) Proteins are first separated by 1D (example shown) or 2D SDS-PAGE. The protein of interest is cut out of the gel and subjected to an in-gel digestion to generate peptides. (**C**) Schematic presentation of a TAP-tagged protein [22]. (**D**) Interacting partners of the target protein (depicted in green and red) are identified from non-specific contaminating proteins (depicted in grey) after two affinity purification steps. The first step is based on affinity of protein A to IgG beads. After removal of unbound proteins, the target protein is released upon TEV cleavage. The second step is based on the affinity of the calmodulin binding peptide for calmodulin beads in the presence of calcium. This purification step removes traces of non-specific contaminating proteins as well as TEV protease. The target protein and its interacting partners are collected from the beads upon EGTA elution and identified by MS.

cleavage with tev protease. Thus the IgG binding domain remains attached to the beads. The second purification step is based on affinity of a calmodulin binding peptide (the second tag) to calmodulin beads in a presence of calcium. Upon removal of any carry-over non-specific proteins and tev protease by washing steps, the over-expressed protein and its interacting partners are eluted from the beads in the presence of chelating agents, such as EGTA, and then separated by SDS-PAGE (Figure 4D). The protocols described so far allow for the identification of proteins as cellular components and members of protein-protein interaction networks. However, if the biological questions are rephrased to monitor differences in protein abundance or their post-translational modifications under various conditions, labelling protocols that allow quantification and direct comparison of two different states in the same analysis are required.

Mass spectrometry as a quantitative tool

Mass spectrometry is not inherently a quantitative technique due to the fact that different molecular species (*e.g.* different peptide sequences) will have diverse ionization efficiencies. Several quantitation methods for MS-based approaches have been devised and can be divided into two types, absolute and relative [14]. The general principle behind these methods takes advantage of the fact that chemical and physical properties of isotopically labelled compounds are identical to their naturally occurring counterparts, with the exception of their mass [23, 24]. Thus stable isotopic nuclei (carbon-13, nitrogen-15, oxygen-18 and deuterium) will not alter ionization efficiency and can be either incorporated into peptides and function as relative references, or included into samples as internal standards [25].

The relative labelling can be either accomplished by metabolic incorporation (*i.e.* stable isotope labelling by amino acids in cell culture, SILAC [23]), by post-harvest chemical derivatization method (*i.e.* isotope-coded affinity tag, ICAT, [24], isobaric tags for relative and absolute quantification, iTRAQ, [26]), or by incubation with $H_2{}^{18}O$ during enzymatic digestion [27].

Absolute quantification is based on 'spiking' known quantities of chemically synthesized stable isotope labelled peptide as internal standards (absolute quantitation, termed AQUA, [28]). The major limitation of the AQUA method is the requirement for chemical synthesis of isotope containing peptide equivalents to proteolytic digest of all generated peptides, and thus is restricted to a small number of pre-selected proteins [14]. With regards to the relative quantification strategies, the least commonly utilized approach is based on proteolytic labelling with oxygen-18 during proteolytic digestion. As this can result in a 2 or 4 Da shift, due to incorporation of either one or two oxygen-18 atoms into the carboxyl terminus of each generated peptide, this may complicate the quantification. Also, if only one oxygen is replaced it might cause an overlap of isotopic envelopes [14, 29], as each MS-detected signal corresponding to a peptide consists of isotope cluster of peaks [7]. These peaks are separated by 1 Da, caused by the 1% probability in nature of each carbon-12 atom to be a carbon-13 isotope [7]. In addition, further complications exist in that variability is added as a result of slow or incomplete incorporation of the oxygen-18 for highly acidic peptides [30].

Mass spectrometry and isotopic in vivo labelling

Biologists often ask whether a certain protein is present within a particular sample, and if the answer is yes, a typical follow-up question is in what amount. In order

to address the quantitative part of their questions, so called "metabolic" or "*in vivo*" labelling methods were developed, which are based on the incorporation of stable isotopes containing essential amino acids into each synthesized protein. The original proof-of-principle SILAC report used incorporation of an isotopically labelled deuterated form of leucine ($^{2}H_3$-Leu) in sample A ('heavy') which was compared to sample B, in which cells were grown in a media containing normal leucine (H-Leu, 'light') [23]. After experimental treatments (in this case myoblast differentiation), the protein populations were harvested and mixed together, separated on SDS-PAGE and proteolytically digested (Figure 5A). As the exact ratio between labelled and unlabelled proteins/peptides is not affected, quantitation (based on peak intensity) can be performed either on the peptide mass or fragment spectrum as the fragmentation patterns are identical and no complications are introduced due to label (3 Da difference). The advantages of this labelling technique include the ability to label many proteins in a cell (as more than half of tryptic peptides contain leucine), there is no difference in labelling efficiencies between samples (cells are exposed to isotope containing amino acids for a prolonged period of time, five days in culture, prior to treatment), and the labelling is consistent among numerous peptides from the same protein and is sequence specific [23]. However, problems were encountered upon reverse phase separation of peptides due to both, the deuterium label and the fact that not all peptides generated by trypsin digest were labelled. In the second generation of labelled amino acids, the stable isotopes were carbon-13 and nitrogen-15, and the label was incorporated into lysine and arginine, instead of leucine. As trypsin (peptide bond cleavage at arginine and lysine residues) is the most commonly used protease in MS-based experiments, when isotopicaly labelled lysine ($^{13}C_6{}^{15}N_2$-Lys) and arginine ($^{13}C_6{}^{15}N_4$-Arg) are incorporated, all detected peptides

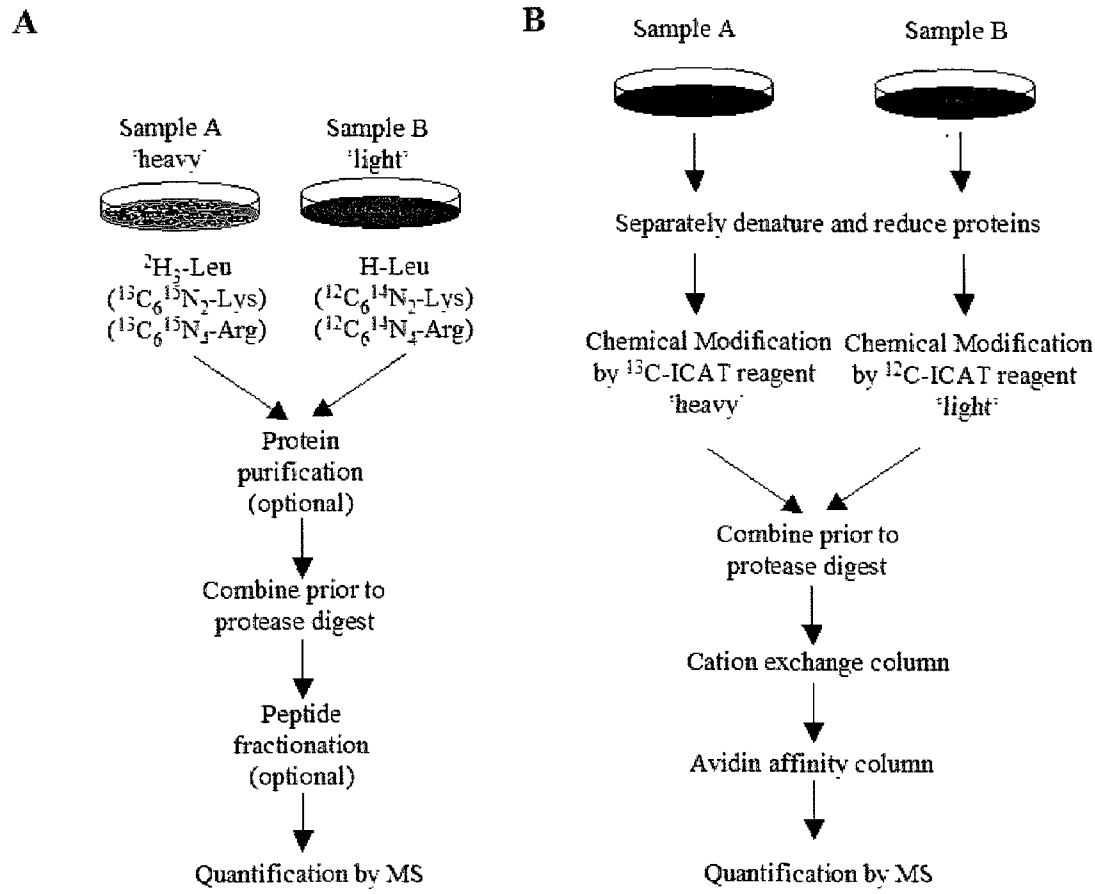


Figure 5. Quantitative mass spectrometry

(**A**) A schematic presentation of the SILAC method, the "metabolic" or "*in vivo*" labelling technique [23]. For sample A, cells are grown in the presence of an isotopically labelled 'heavy' amino acid (either Leu, Lys or Arg, or a combination of Lys and Arg) prior to the experiment. Thus all proteins are isotopically labelled. For sample B, the cells are grown in normal media and thus the proteins will contain 'light' labelled amino acids. One culture is exposed to the desired treatment after which cells are lysed and proteins are extracted from both samples, afterwards both samples are mixed together and might or might not be subjected to purification protocols prior to enzymatic cleavage into peptides. The peptides corresponding to the same protein from two different samples have the same properties, with the exception of the isotopic label (and subsequently different m/z) thus allowing quantification. This therefore allows the determination of the affect of the treatment on protein levels. (**B**) A schematic representation of the ICAT method, one of the "*in vitro*" labelling techniques [24]. ICAT allows isotopic labelling of cysteine residues from numerous samples, including non-living cells and tissues. The first step requires denaturation of proteins followed by treatment with reducing agent, thus generating free thiol residues susceptible to ICAT derivatization. Then the samples under comparative study are combined and subjected to enzymatic proteolysis. The labelled peptides are purified by cation exchange and avidin affinity columns. After biotin cleavage, quantification of purified cysteine containing peptides is performed by MS.

are labelled with an 8 Da mass shift difference for Lys and a 10 Da for Arg containing peptides [31].

An alternative approach to the TAP-tag system for identification of protein-protein interactions involves the use of untagged endogenous proteins, at their normal cellular levels, in a quantitative immunoprecipitation combined with a knockdown screen [32]. This method, QUICK is a hybrid of SILAC, RNA interference (RNAi) and coimmunoprecipitation [32]. As RNAi allows specific depletion of a protein, its interacting partners can be identified upon immunoprecipitation with endogenous antibody, elution, digestion and MS-based comparison of peptide ratios between 'light' cells, where the protein under study has been down-regulated and non-treated 'heavy' cells. After the immunoprecipitation and digestion, MS-analysis detects peptides at a higher intensity corresponding to the protein under study and its interacting partners in the 'heavy' labelled sample (non-treated), as these are decreased or absent from RNAi treated cells; whereas peptides corresponding to contaminating proteins are observed at similar levels [32].

Another promising adaptation of SILAC is in a screen for identification of novel DNA-protein interactions [33]. Protein levels are often a reflection of the regulation of transcription as the case of the presence/absence of specific transcription factors recognition sites within promoter regions (upstream of transcription start site). DNA methylation also affects transcript levels. For example, in mammals in approximately 80% of all genes the promoter regions contain stretches of cytosine-guanine (CG) nucleotides, so called CpG islands [34]. The cytosine within CpG dinucleotides is susceptible to methylation, an event that potentially may disrupt the ability of a transcription factor to bind their recognition sequence. CpG island methylation within

promoter regions of numerous genes have also been shown to display tissue specific patterns as well as a correlation with many malignancies [34]. In order to identify signalling networks mis-regulated due to methylation of a specific CpG island, biotinylated double stranded synthetic oligonucleotide containing either fully methylated cytosine residues within a CpG island or a corresponding non-methylated sequence were separately immobilized onto streptavidin beads [33]. Affinity chromatography was performed separately for DNA containing the methylated CpG island and 'heavy' ($^2H_4$-Lys) labelled nuclear extract and for DNA containing the non-methylated CpG island and 'light' labelled (H-Lys) nuclear extract. Afterwards samples are combined and DNA-protein complexes were eluted by restriction enzyme digestion, as the nucleotides were engineered to contain endonuclease recognition site. Thereafter protein complexes are either immediately digested with a protease in-solution or proteins can be first separated on SDS-PAGE and then in-gel digested. Lysine containing tryptic peptides originating from DNA methylation specific interacting proteins were detected at a higher peak intensity for the $^2H_4$–form (mass shift difference of 4 Da), while DNA affinity purified non-specific proteins displayed 1:1 ratio between isotopic forms [33].

Mass spectrometry and non-isotopic quantitation

In the case where samples have not been or cannot be isotopically labelled, approaches for label-free quantitative proteomic have been developed. These methods allow for extraction of quantitative data from mass spectra, and require implementation of computational algorithms. One of the label-free quantitative methods is based on normalization of the number of observed peptides to the number of observable peptides for the protein under study, the so-called protein abundance index (PAI) [35, 36]. This

was later modified to exponentially modified PAI (emPIA) method, as it was observed, that the number of peptides observed has a logarithmic relationship to the amount of protein within that sample [37]. Furthermore, this method was recently further enhanced by the inclusion of correction factors that allow the measurement of absolute protein concentration per cell from the proportionality between protein abundance and the number of observed peptides [38]. The second label-free quantitative method is only applicable in highly reproducible liquid-chromatography interfaced MS-instruments. This method is based on extracted ion current, which is the area under the curve, when signal intensity as peptide elutes from chromatographic column, is plotted versus time [39]. The extracted ion current is linearly related to the amount of particular peptide (extracted ion currents for different peptides of the same protein are very different) under the same experimental conditions [14]. Samples under comparative analysis are processed and analyzed separately, one after another. It is noted that slight changes in sample processing or the presence of contaminants, such detergents, interfere with quantification [14]. Applications of this label-free quantitative analysis include, but are not limited to, protein correlation profiling, an adaptation to provide protein abundance profiles upon fractionation procedures [40] and detection of changes in the cell surface proteome [41].

Mass spectrometry and isotopic in vitro labelling

A second category of isotopic labelling strategies is based on chemically targeting reactive groups on side chains of amino acids or peptide termini. Some of these labelling approaches are performed at the peptide level and some at the protein level, depending upon the experimental questions. As the above-described *in vivo* labelling techniques are limited to cells that can be cultured in appropriate media, the alternative approaches for

incorporation of labels allowing quantitative proteome analysis can be performed on a wide variety of samples, including patient specimens.

The most popular of the isobaric mass tags, iTRAQ allows concurrent analysis of up to eight different samples and consists of an amine-specific reactive group, a reporter mass group (variable mass of 114-117 Da (for 4-plex assays) or 113-121 Da (for 8-plex assays)) and a balancer group (variable mass of 191-188 Da or 192-184 Da) [26]. The iTRAQ label introduces a highly basic group at the N-terminus of peptides and lysine side chains, and during MS/MS the reporter groups are released as singly charged ions of masses at 114, 115, 116 and 117 m/z (in a 4-plex experiment) thus allowing quantification based on their ratios.

Recently introduced (by Applied Biosystems) the non-isobaric, amine based isotope-coded reagent, mTRAQ, allows concurrent quantification of three samples, as addition of ^{13}C and ^{15}N containing tags as precursors that result in a mass shift of 0, 4 and 8 Da for arginine and 0, 8 and 16 Da for lysine tryptic peptides [29].

Another class of labelling reagents has specificity towards sulfhydryl groups, a reduced form of the side-chain in cysteine residues. These isotope-coded affinity tags (ICAT) contain a thiol-specific reactive group (iodoacetamide) and a biotin affinity tag, separated by an isotopically coded linker [24]. This quantitative labelling technique was designed in order to simplify the complex peptide pool prior to MS-analysis, while maintaining high proteome coverage. Cysteine was chosen because it is a rare, but functionally essential amino acid, with most proteins containing at least one cysteine. This fact reduces the complexity of the peptide sample by approximately an order of

magnitude (Figure 5B) [24]. The major caveat of this method is that the representation of proteins that lack cysteine residue is compromised. Based on current estimates this represents approximately 5% of human proteins [42]. In the first generation of ICAT reagents the difference between 'light' and 'heavy' was 8 Da, due to incorporation of eight deuterium atoms into the 'heavy' tag [24]. There were a few shortcomings related to the design of these labelling tags, such as a shift in the elution of corresponding peptides from reverse phase chromatography due to deuterium modification and ambiguities in assigning peptides with more than one cysteine. Note, a difference of 16 Da may represent the presence of either two labelled cysteine residues or oxidation of methionine. An additional difficulty was the possibility of undesired fragmentation at the biotin tag, instead of fragmentation at the peptide backbone during MS/MS. This led to development of a second generation of ICAT reagents [43], which are now commercially available. These ICAT reagents include an acid-cleavable linker group between the biotin moiety and sulfhydryl reactive isotope tag, while the 'heavy' reagent contains nine ^{13}C atoms instead of eight deuterium atoms [43, 44]. These reagents allow the comparison of the content and relative abundance between two samples, by first reducing the side chains of cysteinyl residues in a sample representing one cell state followed by derivatization with 'light' ICAT reagent. In the second sample, equivalent groups are first reduced and then derivatized with 'heavy' ICAT reagent [24]. After labelling, the two samples were combined and proteolytically digested; thereafter the resulting peptides were separated by multidimensional chromatography, including isolation of tagged peptides by avidin-affinity. Quantification is based on comparison of the relative signal intensities (ratios) between peptide ions pairs of identical sequence, with a mass difference due to incorporation of 'light' and 'heavy' ICAT tags which are chemically identical, thus

allowing an accurate measure of the relative abundance of the peptide (hence the protein) [24]. The application of ICAT-labelling to study changes in protein abundance at the entire proteome level ranges from analysis of differences in microsomal fractions between naïve and differentiated human myeloid leukemia cells [45], in tracheal tissue of cystic fibrosis patients [43], and serum of hepatocellular carcinoma patients [46], just to mention a few from a long list.

The thiol group of cysteine residue is very susceptible to oxidation by reactive oxygen species or reactive nitrogen species [47]. Furthermore, thiol groups (-SH) are essential for the function of many proteins, in that oxidation (in a form of disulfide, -S-S) may increase or decrease a protein's activity, thus affecting their function. ICAT reagents have provided an invaluable tool to identify which cysteines are susceptible to oxidation (for example after exposure to hydrogen peroxide) and for the quantification of the amount of oxidized cysteine residues [48, 49]. This procedure is based on the premise that oxidized cysteine residues are not susceptible to ICAT derivatization, when samples are labelled under non-reducing conditions. Thus normal sample would be labelled with 'light' ICAT reagent, and the experimental sample exposed to oxidant stress, would be labelled with 'heavy' ICAT reagent, thus the decrease in specific labelling of oxidant-sensitive cysteine thiols can be quantified from the ratio of ICAT reagents.

One of key redox modulators that contain a CysXXCys motif within two active sites is protein disulfide isomerase (PDI). In collaboration with Dr. B. Mutus, Department of Chemistry and Biochemistry, University of Windsor, ICAT labelling was utilized to identify and quantitate redox sensitive cysteines on PDI after auto-oxidation. This study is presented in detail in Chapter 3.

In addition to identification of redox-sensitive thiols, the ICAT in combination with labelling has been adapted to trap differential thiols, in a method called OxICAT [50]. The OxICAT method is based on sequential labelling of the same sample; first free thiols are labelled with 'light' ICAT reagent after protein(s) denaturation and afterwards disulfides are disrupted by strong reducing agent, thus allowing for these thiols to be labelled with 'heavy' ICAT reagent. Cysteines that were originally present as thiols, will be detected only as 'light' ICAT labelled; whereas cysteines which were originally oxidized, will be identified by the 'heavy' ICAT modification. This will result in a 9 Da shift if the peptide contains one cysteine and if more cysteines, multiples thereof. If cysteines are at redox equilibrium, then both forms ('light' and 'heavy' ICAT labelled) for the same peptide would be detected. This approach also provides a novel method to monitor the kinetics of thiol oxidation by trapping oxidation intermediates within a purified protein (for example upon oxidative stress, thus activating or inactivating the protein) [50].

Thiols can also be modified by either S-nitrosylation or S-glutathiolation [51]. The MS-based approach to identify Cys-SNO modifications, SNO-Site Identification method SNOSID, is essentially an adaptation of biotin-switch assay [52, 53]. First, all free thiols in the protein mixture are methylated, then Cys-SNO are selectively reduced (by ascorbate), thus generating a new unmodified thiols, at which position biotin will be attached through a disulfide bond (upon addition of N-[6-(biotinamido)hexyl]-3'-(2'-pyridyldithio)propionamide, biotin-HPDP). At this time, proteins are proteolytically digested, biotin-tagged peptides are isolated by avidin-affinity, and subsequently released by the reduction of the disulfide linker. As this last described method is not based on

ICAT technology, it serves as an example that alternative methods are available to study cysteine modifications.

Mass spectrometry and post-translational modifications

So far, MS-based approaches designed to identify novel proteins and quantitative differences among samples have been described. Nevertheless, biological functions of most proteins are regulated through post-translational modifications. The ability to unambiguously identify these post-secondary modifications (and their location), further highlights the power of MS. Identification is achieved from MS/MS spectra analysis, as the observed fragmentation pattern is altered due to the mass shift introduced by the modification. The acquired spectra can be compared to a predicted fragmentation pattern (generated by open-source software) including defined modifications introduced at a specific amino acid. For example, acetylation on lysine residue introduces a mass shift difference of +42 Da of the parent ion $[MH]^+$ compared to unmodified peptide in MS spectra (prior to fragmentation), while the position of modified lysine can be assigned by +42 Da shift detected in MS/MS spectra at the position of this residue. In the case of methylation, the extent of mass shifts introduces depends on the number of methyl groups present, *i.e.* +14 Da for methylation, +28 Da for dimethylation and +42 Da for trimethylation. Ubiquitination plays an important role in protein stability, as ubiquitinated proteins are targeted for degradation [54]. The MS-identification of lysine residues originally modified by ubiquitination is based on detection of ubiquitin remnants, in a form of a covalently attached glycine-glycine dipeptide, the signature peptide (mass shift difference at the lysine residue of 114.1 Da), after trypsin proteolysis; sustained by trypsin missed cleavage at this lysine residue (proteolysis does not occur at modified

lysines) [55]. Protein modifications by carbohydrates (O-linked and N-linked glycosylation) have been increasingly recognized as biochemical alterations associated with malignant transformations and tumorigenesis [56]. Due to the incredible variability in the glycan structures, and complexity of these modifications, methodologies applied to their identification and characterization are beyond the scope of this overview.

Mass spectrometry and identification of phosphorylation sites

Last, but not least, phosphorylation is one of the most important regulatory modifications, affecting a broad range of biological functions and processes. Phosphorylation occurs most commonly on serine residues (86.4%), followed by threonine residues (11.8%) and rarely on tyrosine residues (1.8%) (values are based on *in vivo* phosphoproteome analysis) [57]. Other amino acids that can be phosphorylated are histidine and glutamic acid [42]. Phosphorylation is detected as a modification of the peptide by a +80 Da shift in the MS spectra in comparison to the non-phosphorylated peptide and typical dominant –98 Da and –80 Da fragments detected in the MS/MS spectra due to neutral losses (the mass of H_3PO_4, and HPO_3 respectively, derived from phosphoserine and phoshothreonine). In the case of phosphotyrosine, a specific immonium ion (an internal fragment containing just a single side chain of an amino acid) is detected at 216.02 m/z [58].

Due to the dynamic nature of phosphorylation, phosphorylated proteins are often present at low levels, thus the fraction of phosphorylated peptides will be at a much lower concentration compared to their non-phosphorylated counter-parts (high abundance peptides cause ion suppression during ionization process) [59]. In addition to low stoichiometry, phosphopeptides also suffer from low ionization efficiency [59]. These

problems can be circumvented by techniques that promote enrichment of phosphopeptides.

Originally, three methods were described for the identification of phosphopeptides from a complex protein mixture. The first one included replacement of phosphate groups by ethanedithiol, followed by the addition of a biotin tag, the latter allowing phosphopeptide affinity purification [60]. A limitation of this method includes complications that arise during the derivatization process and the inability to detect phosphotyrosine residues, as only phosphoserine and phosphothreonine undergo β-elimination [61]. The second method was based on a chemistry intensive six-step derivatization/purification protocol for tryptic peptides, where the capture of phosphopeptides was accomplished through a reversible covalent linkage to a solid support [62]. The third method employs phosphopeptide enrichment from whole cell lysates by immobilized metal affinity chromatography (IMAC) on peptides derivatized by esterification of carboxylate groups [63]. IMAC has become one of the most commonly used strategies for phosphopeptide enrichment [5] and will be described in further details later.

In addition to IMAC, numerous techniques have recently been developed to enrich for phosphopeptides. For example, strong cation exchange chromatography (SCX-based strategy), which separates peptides according to ionic charge, allows enrichment of tryptic phosphopeptides from unmodified peptides (singly charged species elute before multiply charges ones) [64]. At low pH (2.7) only lysine, arginine, histidine and the amino terminus are charged with a net charge state of +2 on non-phosphorylated peptide, while phosphate groups maintain a negative charge with a net charge state of +1 on a

phosphopeptide [64]. The major caveats of this enrichment technique include the fact that it is limited to studying trypsin generated proteolytic fragments as well as having an inability to detect phosphopeptides containing a single histidine residue [65]. The presence of histidine within a phosphopeptide alters the net charge to +2, the same as non-phosphopeptides, thus effecting the elution properties of this phosphopeptide from the SCX-column and hindering detection [65]. Nevertheless, this phospho-enrichment protocol is powerful and has allowed the phosphoproteomic analysis of the developing mouse brain [65]. Furthermore, in combination with other phosphopeptide enrichment methods, SCX-based strategies have been implemented as the first fractionation step in numerous studies, including the effect of epidermal growth factor induced simulation in HeLa cells [57], profiling of mitotic phosphorylation [66] and analyses of the phosphoproteome in mouse liver [67], and *Drosophila melanogaster* embryos [68].

A motif-specific phosphopeptide enrichment method has also been described that is useful for characterizing specific kinase targets. The method is based on the precipitation of phosphopeptides with Ba^{2+} in the presence of acetone under varying pH [69]. For example, Ba^{2+}-based precipitation in combination with pH dependent fractionation and MudPIT separation of peptides was used to analyze nuclear extracts from HeLa cells. When the identified phosphopeptides were analyzed in a pH dependent manner, at pH 3.5, 86% of the phosphopeptides were unique and the sequence adjacent to the phosphorylated residue contained acidic amino acids; at pH 4.6 70% and at pH 8.0 56% of the phosphopeptides were unique. Ba^{2+}-based precipitation at pH 4.6 and 8.0 displayed an increased preference for proline residues C-terminal to the phosphorylation site. Overall, the pattern of Ba^{2+} precipitation of phosphopeptides at each pH coincided with consensus sequences for specific kinases [69].

As phosphorylation at tyrosine residue ranges between 0.05% (as determined by phosphoamino acid analysis [70]) to 1.8% (as determined by phosphoproteome analysis [57]) of the total cellular phosphorylation, selective enrichment for phosphotyrosines containing phosphopeptides is necessary when the quest is to identify tyrosine phosphorylation sites. The method of choice to selectively enrich for these phosphopeptides is based on peptide affinity purification with an anti-phosphotyrosine antibody. The immuno-affinity phosphopeptide strategy was first applied to profile analysis of oncogenic Src (tyrosine kinase) down-stream targets and comparative analysis of two cancer cell lines derived from anaplastic large cell lymphomas [71]. The same experimental approach was utilized to study the phosphotyrosine proteome in mouse brain [72].

In addition to IMAC, a metal oxide affinity chromatography (MOAC) based method, which utilizes ZrO_2 and TiO_2, has been shown to selectively bind phosphopeptides [73-75]. The limitation of this enrichment method is similar to IMAC (discussed below), *i.e.* the possibility of retention of acidic non-phosphopeptides [74, 75]. When a phosphopeptide standard mixture was isotopically labelled with iTRAQ (to allow direct comparison) and phosphopeptides were enriched by either TiO_2 or Fe^{3+}-based affinity and then mixed together prior to MS-analysis, comparable performance between these two technique was observed [58]. Furthermore, when phosphopeptides from biological sample were SILAC labelled (to quantitate changes in phosphorylation after EGF stimulation), phosphotyrosine containing proteins were immuno-precipitated with an anti-phosphotyrosine antibody, protein complexes were separated on a gel and after in-gel digestion, peptides were enriched by either TiO_2 or Fe^3, results were similar [58].

Immobilized metal ions, such as Fe^{3+} and Al^{3+} bind phosphoproteins and phosphopeptides with high specificity [76, 77]. Furthermore, after enrichment of phosphopeptides on either Fe^{3+} and Ga^{3+} loaded agarose, phosphorylated residues can be identified by MS [78]. This experimental protocol is the basis of commercially available PHOS-Select resin (from Sigma-Aldrich). The major limitations of this enrichment technique include binding of non-specific peptides and variability in the retention of phosphopeptides [63, 79]. IMAC enrichment of phosphopeptides has been employed to identify human mitotic spindle phosphoproteins [80], in the phosphoproteome analysis of α-factor arrested *Saccharomyces cerevisiae* [81] and phosphoproteins in *Schizosaccharomyces pombe* [82].

Mass spectrometry and dual specificity phosphatases

Dual specificity phosphatases (DUSPs) are able to remove a phosphate group from phosphoserine, phosphothreonine and phosphotyrosine residues [83, 84]. This very heterogenous sub-group is a subfamily of the protein tyrosine phosphatase (PTP) superfamily [83]. All PTP family members (including DUSPs) contain the invariant catalytic sequence $C(X)_5R$ and the hydrolysis of phosphorylated substrate involves formation of a stable phosphoryl-intermediate [85, 86]. The DUSPs are the most diverse group in terms of substrate specificity, and include the following families; MKPs (mitogen-activated protein kinase phosphatases), Slingshots, PRLs (phosphatase of regenerating liver), Cdc14s, PTENs (phosphatase and tensin homologue deleted on chromosome 10), MTMRs (myotubularins) and the atypical DUSPs [83, 84].

The atypical DUSPs, a group of 16 members, are generally small, under 250 amino acids, and lack specific mitogen activated protein kinase (MAPK) targeting motifs even through a few family members can dephosphorylate at least one of the MAPKs (either Erk, Jnk or p38). These includes DUSP3 (VHR) [87], DUSP14 [88], DUSP18 [89], DUSP19 [90], DUSP22 [91], DUSP23 [92] and DUSP26 [93]. Other members display a broad range of substrate specificities. For example, laforin dephosphorylates complex carbohydrates [94] and DUSP11 interacts with RNA [95]. If all of this heterogeneity and variability within this sub-class of DUSPs is taken into consideration, we can expect that hYVH1 (also known as DUSP12), the DUSP of interest in the Vacratsis lab and the only atypical DUSP that contains a zinc-binding domain, to have a unique role.

YVH1 orthologues are extremely conserved throughout evolution; present in species as diverse as *Plasmodium falciparum* [96] to humans signifying YVH1 orthologues as regulators of conserved cellular functions. Interestingly, YVH1 orthologues are present as single gene copies even in advanced mammals such as humans, suggesting that there has been evolutionary pressure against *yvh1* gene duplication or deletion.

Recent projects from our laboratory have used several MS-based approaches to discover novel hYVH1 functions. A MS-based approach was utilized to identify the cytoprotective heat shock protein 70 (Hsp70) as the first known hYVH1 interacting protein [97]. Phosphatase assays showed that hYVH1 is a heat stable enzyme whose activity during oxidative conditions can be modestly increased in the presence of Hsp70. Furthermore, this result led to the discovery that hYVH1 is a novel cell survival

phosphatase and can prevent both thermal and oxidative stress induced cell death. This cytoprotective property requires catalytic competence and the zinc binding domain [97].

Our laboratory also used differential thiol labelling and MS to reveal that the cysteines in the zinc binding domain serve a redox buffer role, allowing hYVH1 to function under oxidative conditions known to inactivate cysteine based phosphatases [98]. In the present study, outlined in Chapter 4, a MS-based approach was also utilized to identify phosphorylation sites within hYVH1. Not only are these the first phosphorylation sites reported for any YVH1 orthologues, but this result led us to discover that hYVH1 expression affects cell cycle progression.

Objectives of this Thesis

Recent developments in the mass spectrometry field have facilitated the applicability of this experimental tool to study a variety of biological questions. These include, but are not limited to, identification of proteins or post-translational modifications, and quantitative analysis of proteins after modification of peptides with isotopically labelled tags. This dissertation is comprised of three research chapters that examine various applications of MS-based methods to study biological questions.

Chapter 2 investigates the identity of a 21 kDa protein from encysted embryos of brine shrimp *Artemia franciscana* as a group 1 Late embryogenesis abundant protein by a MS-based *de novo* sequencing approach.

Chapter 3 examines the redox status of cysteine residues within recombinant human protein disulfide isomerase after auto-oxidation conditions employing the isotope-coded affinity tag (ICAT) technology and mass spectrometry.

Chapter 4 identifies three *in vivo* phosphorylation sites on hYVH1 using a combination of a phosphopeptide enrichment protocol and mass spectrometry. This study also examines the effect of overexpressing hYVH1 and phosphospecific mutants on hYVH1 localization and progression through the cell cycle. In addition, phenotypes observed due to down-regulation of hYVH1 with siRNA are also described.

References

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

2. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

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

4. Yates, J.R., 3rd, *Mass spectral analysis in proteomics.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 297-316.

5. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications.* Annu Rev Biomed Eng, 2009. **11**: p. 49-79.

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

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

8. Han, X., A. Aslanian, and J.R. Yates, 3rd, *Mass spectrometry for proteomics.* Curr Opin Chem Biol, 2008. **12**(5): p. 483-90.

9. Gevaert, K., et al., *Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry.* Electrophoresis, 2001. **22**(9): p. 1645-51.

10. Shukla, A.K. and J.H. Futrell, *Tandem mass spectrometry: dissociation of ions by collisional activation.* J Mass Spectrom, 2000. **35**(9): p. 1069-90.

11. Link, A.J., et al., *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.

12. Nissan, T.A., et al., *60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm.* Embo J, 2002. **21**(20): p. 5539-47.

13. Foster, L.J., et al., *A mammalian organelle map by protein correlation profiling.* Cell, 2006. **125**(1): p. 187-99.

14. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol, 2005. **1**(5): p. 252-62.

15. Picotti, P., et al., *Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics.* Cell, 2009. **138**(4): p. 795-806.

16. Lange, V., et al., *Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring.* Mol Cell Proteomics, 2008. **7**(8): p. 1489-500.

17. Anderson, L. and C.L. Hunter, *Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins.* Mol Cell Proteomics, 2006. **5**(4): p. 573-88.

18. Bertsch, A., et al., *Optimal de novo Design of MRM Experiments for Rapid Assay Development in Targeted Proteomics.* J Proteome Res, 2010. **9**(5): p. 2696-704

19. Keough, T., R.S. Youngquist, and M.P. Lacey, *A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry.* Proc Natl Acad Sci U S A, 1999. **96**(13): p. 7131-6.

20. Keough, T., M.P. Lacey, and R.S. Youngquist, *Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry.* Rapid Commun Mass Spectrom, 2000. **14**(24): p. 2348-56.

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

22. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.

23. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

24. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.

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

26. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

27. Schnolzer, M., P. Jedrzejewski, and W.D. Lehmann, *Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein*

sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. Electrophoresis, 1996. **17**(5): p. 945-53.

28. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.

29. Elliott, M.H., et al., *Current trends in quantitative proteomics.* J Mass Spectrom, 2009. **44**(12): p. 1637-60.

30. Gevaert, K., et al., *Stable isotopic labeling in proteomics.* Proteomics, 2008. **8**(23-24): p. 4873-85.

31. Gruhler, A., et al., *Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway.* Mol Cell Proteomics, 2005. **4**(3): p. 310-27.

32. Selbach, M. and M. Mann, *Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK).* Nat Methods, 2006. **3**(12): p. 981-3.

33. Mittler, G., F. Butter, and M. Mann, *A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements.* Genome Res, 2009. **19**(2): p. 284-93.

34. Esteller, M., *Cancer epigenomics: DNA methylomes and histone-modification maps.* Nat Rev Genet, 2007. **8**(4): p. 286-98.

35. Rappsilber, J., et al., *Large-scale proteomic analysis of the human spliceosome.* Genome Res, 2002. **12**(8): p. 1231-45.

36. Sanders, S.L., et al., *Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry.* Mol Cell Biol, 2002. **22**(13): p. 4723-38.

37. Ishihama, Y., et al., *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics, 2005. **4**(9): p. 1265-72.

38. Lu, P., et al., *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.* Nat Biotechnol, 2007. **25**(1): p. 117-24.

39. Chelius, D. and P.V. Bondarenko, *Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry.* J Proteome Res, 2002. **1**(4): p. 317-23.

40. Andersen, J.S., et al., *Proteomic characterization of the human centrosome by protein correlation profiling.* Nature, 2003. **426**(6966): p. 570-4.

41. Schiess, R., et al., *Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry.* Mol Cell Proteomics, 2009. **8**(4): p. 624-38.

42. Zhang, H., W. Yan, and R. Aebersold, *Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes.* Curr Opin Chem Biol, 2004. **8**(1): p. 66-75.

43. Hansen, K.C., et al., *Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography.* Mol Cell Proteomics, 2003. **2**(5): p. 299-314.

44. Yi, E.C., et al., *Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme.* Proteomics, 2005. **5**(2): p. 380-7.

45. Han, D.K., et al., *Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry.* Nat Biotechnol, 2001. **19**(10): p. 946-51.

46. Kang, X., et al., *Serum protein biomarkers screening in HCC patients with liver cirrhosis by ICAT-LC-MS/MS.* J Cancer Res Clin Oncol, 2010.

47. Forman, H.J., et al., *The chemistry of cell signaling by reactive oxygen and nitrogen species and 4-hydroxynonenal.* Arch Biochem Biophys, 2008. **477**(2): p. 183-95.

48. Sethuraman, M., et al., *Isotope-coded affinity tag (ICAT) approach to redox proteomics: identification and quantitation of oxidant-sensitive cysteine thiols in complex protein mixtures.* J Proteome Res, 2004. **3**(6): p. 1228-33.

49. Sethuraman, M., et al., *Isotope-coded affinity tag approach to identify and quantify oxidant-sensitive protein thiols.* Mol Cell Proteomics, 2004. **3**(3): p. 273-8.

50. Leichert, L.I., et al., *Quantifying changes in the thiol redox proteome upon oxidative stress in vivo.* Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8197-202.

51. Mallis, R.J., J.E. Buss, and J.A. Thomas, *Oxidative modification of H-ras: S-thiolation and S-nitrosylation of reactive cysteines.* Biochem J, 2001. **355**(Pt 1): p. 145-53.

52. Jaffrey, S.R., et al., *Protein S-nitrosylation: a physiological signal for neuronal nitric oxide.* Nat Cell Biol, 2001. **3**(2): p. 193-7.

53. Hao, G., et al., *SNOSID, a proteomic method for identification of cysteine S-nitrosylation sites in complex protein mixtures.* Proc Natl Acad Sci U S A, 2006. **103**(4): p. 1012-7.

54. Ande, S.R., J. Chen, and S. Maddika, *The ubiquitin pathway: an emerging drug target in cancer therapy.* Eur J Pharmacol, 2009. **625**(1-3): p. 199-205.

55. Peng, J., et al., *A proteomics approach to understanding protein ubiquitination.* Nat Biotechnol, 2003. **21**(8): p. 921-6.

56. Reis, C.A., et al., *Alterations in glycosylation as biomarkers for cancer detection.* J Clin Pathol, 2010. **63**(4): p. 322-9.

57. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.

58. Liang, X., et al., *Quantitative comparison of IMAC and TiO2 surfaces used in the study of regulated, dynamic protein phosphorylation.* J Am Soc Mass Spectrom, 2007. **18**(11): p. 1932-44.

59. Liao, P.C., et al., *An approach to locate phosphorylation sites in a phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry.* Anal Biochem, 1994. **219**(1): p. 9-20.

60. Oda, Y., T. Nagasu, and B.T. Chait, *Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome.* Nat Biotechnol, 2001. **19**(4): p. 379-82.

61. Thompson, A.J., et al., *Characterization of protein phosphorylation by mass spectrometry using immobilized metal ion affinity chromatography with on-resin beta-elimination and Michael addition.* Anal Chem, 2003. **75**(13): p. 3232-43.

62. Zhou, H., J.D. Watts, and R. Aebersold, *A systematic approach to the analysis of protein phosphorylation.* Nat Biotechnol, 2001. **19**(4): p. 375-8.

63. Ficarro, S.B., et al., *Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae.* Nat Biotechnol, 2002. **20**(3): p. 301-5.

64. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins.* Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12130-5.

65. Ballif, B.A., et al., *Phosphoproteomic analysis of the developing mouse brain.* Mol Cell Proteomics, 2004. **3**(11): p. 1093-101.

66. Dephoure, N., et al., *A quantitative atlas of mitotic phosphorylation.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10762-7.

67. Villen, J., et al., *Large-scale phosphorylation analysis of mouse liver.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1488-93.

68. Zhai, B., et al., *Phosphoproteome analysis of Drosophila melanogaster embryos.* J Proteome Res, 2008. **7**(4): p. 1675-82.

69. Ruse, C.I., et al., *Motif-specific sampling of phosphoproteomes.* J Proteome Res, 2008. **7**(5): p. 2140-50.

70. Hunter, T. and B.M. Sefton, *Transforming gene product of Rous sarcoma virus phosphorylates tyrosine.* Proc Natl Acad Sci U S A, 1980. **77**(3): p. 1311-5.

71. Rush, J., et al., *Immunoaffinity profiling of tyrosine phosphorylation in cancer cells.* Nat Biotechnol, 2005. **23**(1): p. 94-101.

72. Ballif, B.A., et al., *Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain.* J Proteome Res, 2008. **7**(1): p. 311-8.

73. Zhou, H., et al., *Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis.* Electrophoresis, 2007. **28**(13): p. 2201-15.

74. Larsen, M.R., et al., *Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns.* Mol Cell Proteomics, 2005. **4**(7): p. 873-86.

75. Pinkse, M.W., et al., *Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. **76**(14): p. 3935-43.

76. Muszynska, G., et al., *Model studies on iron(III) ion affinity chromatography. II. Interaction of immobilized iron(III) ions with phosphorylated amino acids, peptides and proteins.* J Chromatogr, 1992. **604**(1): p. 19-28.

77. Andersson, L. and J. Porath, *Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography.* Anal Biochem, 1986. **154**(1): p. 250-4.

78. Zhou, W., et al., *Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beads by matrix-assisted laser desorption/ionization mass spectrometry.* J Am Soc Mass Spectrom, 2000. **11**(4): p. 273-82.

79. Posewitz, M.C. and P. Tempst, *Immobilized gallium(III) affinity chromatography of phosphopeptides.* Anal Chem, 1999. **71**(14): p. 2883-92.

80. Nousiainen, M., et al., *Phosphoproteome analysis of the human mitotic spindle.* Proc Natl Acad Sci U S A, 2006. **103**(14): p. 5391-6.

81. Li, X., et al., *Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae.* J Proteome Res, 2007. **6**(3): p. 1190-7.

82. Wilson-Grady, J.T., J. Villen, and S.P. Gygi, *Phosphoproteome analysis of fission yeast.* J Proteome Res, 2008. **7**(3): p. 1088-97.

83. Alonso, A., et al., *Protein tyrosine phosphatases in the human genome.* Cell, 2004. **117**(6): p. 699-711.

84. Patterson, K.I., et al., *Dual-specificity phosphatases: critical regulators with diverse cellular targets.* Biochem J, 2009. **418**(3): p. 475-89.

85. Denu, J.M. and J.E. Dixon, *A catalytic mechanism for the dual-specific phosphatases.* Proc Natl Acad Sci U S A, 1995. **92**(13): p. 5910-4.

86. Denu, J.M., et al., *The catalytic role of aspartic acid-92 in a human dual-specific protein-tyrosine-phosphatase.* Biochemistry, 1995. **34**(10): p. 3396-403.

87. Todd, J.L., K.G. Tanner, and J.M. Denu, *Extracellular regulated kinases (ERK) 1 and ERK2 are authentic substrates for the dual-specificity protein-tyrosine phosphatase VHR. A novel role in down-regulating the ERK pathway.* J Biol Chem, 1999. **274**(19): p. 13271-80.

88. Marti, F., et al., *Negative-feedback regulation of CD28 costimulation by a novel mitogen-activated protein kinase phosphatase, MKP6.* J Immunol, 2001. **166**(1): p. 197-206.

89. Wu, Q., et al., *Dual specificity phosphotase 18, interacting with SAPK, dephosphorylates SAPK and inhibits SAPK/JNK signal pathway in vivo.* Front Biosci, 2006. **11**: p. 2714-24.

90. Zama, T., et al., *Scaffold role of a mitogen-activated protein kinase phosphatase, SKRP1, for the JNK signaling pathway.* J Biol Chem, 2002. **277**(26): p. 23919-26.

91. Aoyama, K., et al., *Molecular cloning and characterization of a novel dual specificity phosphatase, LMW-DSP2, that lacks the cdc25 homology domain.* J Biol Chem, 2001. **276**(29): p. 27575-83.

92. Wu, Q., et al., *Molecular cloning and characterization of a novel dual-specificity phosphatase 23 gene from human fetal brain.* Int J Biochem Cell Biol, 2004. **36**(8): p. 1542-53.

93. Hu, Y. and N.F. Mivechi, *Association and regulation of heat shock transcription factor 4b with both extracellular signal-regulated kinase mitogen-activated protein kinase and dual-specificity tyrosine phosphatase DUSP26.* Mol Cell Biol, 2006. **26**(8): p. 3282-94.

94. Worby, C.A., M.S. Gentry, and J.E. Dixon, *Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates.* J Biol Chem, 2006. **281**(41): p. 30412-8.

95. Yuan, Y., D.M. Li, and H. Sun, *PIR1, a novel phosphatase that exhibits high affinity to RNA . ribonucleoprotein complexes.* J Biol Chem, 1998. **273**(32): p. 20347-53.

96. Kumar, R., et al., *A zinc-binding dual-specificity YVH1 phosphatase in the malaria parasite, Plasmodium falciparum, and its interaction with the nuclear protein, pescadillo.* Mol Biochem Parasitol, 2004. **133**(2): p. 297-310.

97. Sharda, P.R., et al., *The dual-specificity phosphatase hYVH1 interacts with Hsp70 and prevents heat-shock-induced cell death.* Biochem J, 2009. **418**(2): p. 391-401.

98. Bonham, C.A. and P.O. Vacratsis, *Redox regulation of the human dual specificity phosphatase YVH1 through disulfide bond formation.* J Biol Chem, 2009. **284**(34): p. 22853-64.

CHAPTER 2

Identification of group 1 LEA protein from *A. franciscana* by *de novo* MS

This chapter incorporates the outcome of joint research undertaken in collaboration with Michelle A. Sharon under the supervision of Dr. Alden H. Warner, Department of Biological Sciences. The collaboration is covered in Chapter 2 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the authors; Michelle purified a 21 kDa protein from *Artemia franciscana*, PCR amplified cDNA from an embryonic library corresponding to its gene and performed the initial characterization of this protein. My contribution to this manuscript was the *de novo* identification of this protein using mass spectrometry (MS). Since the genome of this organism has not yet been sequenced, this included *in silico* analysis for the identification. Based on the peptide sequence corresponding to a conserved motif, the 21 kDa protein was identified as a member of group 1 Late embryogenesis abundant (LEA) family. Once Michelle completed the gene amplification, I was able to match the peptide sequences generated by MS and thus further confirm the identity of the 21 kDa protein.

The results were published: Sharon, M.A., Kozarova, A., Clegg, J.S., Vacratsis, P.O., and Warner, A.H. (2009). Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp *Artemia franciscana*. Biochem Cell Biol *87*, 415-430

Introduction

The brine shrimp, *Artemia franciscana* is a primitive crustacean, able to survive under very harsh hypersaline conditions with their embryos among the most resistant multicellular eukaryotes to environmental stress [1]. Encysted embryos can survive severe desiccation, anoxia, extremes in temperature, and various forms of radiation [1]. Under conditions of stress *Artemia* embryos, which are encysted, undergo a reversible state of dormancy termed diapause that involves both, developmental and metabolic arrest. Previously identified proteins restricted to diapause include p26, a member of the small heat shock-α-crystallin family [2, 3] and artemin, a heat-stable RNA binding protein [4]. However, neither of these proteins has been directly implicated in the ability of these embryos to undergo virtually complete water-loss, thus implying the involvement of another protein(s) in addition to the known function of carbohydrates.

Late embryogenesis abundant proteins (LEA), a group of hydrophilic proteins, have been proposed to protect cellular and molecular structures from low temperature and desiccation-induced damage [5, 6]. Originally identified only in plants, expression of these proteins is pre-programmed during the later stages of seed development. However, they can also be induced by environmental stress, including desiccation, cold and osmotic shock [7]. LEA proteins have been categorized into 6 groups based on their primary amino acid sequence and expression patterns. Group 1-3 LEA proteins are characterized by the presence of specific sequence motifs. The group 1 LEA proteins contain a 20 hydrophilic amino acid motif. The group 2 LEA proteins exhibit at least two of three distinct sequence motifs (termed the Y-, S- and K-segments). The group 3 LEA family members contain multiple copies of an 11 amino acid motif [7]. Categorization within

group 4-7 LEA proteins (with exception of Group 5, that contains hydrophobic or atypical LEA proteins) remains under discussion [7], despite the fact that several distinctive highly conserved motifs have been described [8].

In addition to plants, LEA proteins have also been described in animal species including the nematode *Caenorhabditis elegans* [9]. *C. elegans* is able to enter into a reversible state of dormancy during development, the so called dauer juvenile larval stage, when the food supply is limited. While in the dauer form, these worms exhibit increased tolerance to desiccation and express group 3 LEA-like proteins (*Ce-lea-1*), while the adults are desiccation sensitive [9].

The most complex metazoan for which group 3 LEA proteins have been described thus far is *A. franciscana* [8]. AfrLEA 1 and AfrLEA 2 were identified by a differential developmental stage specific cDNA library screen [10].

In addition, a proteomic study identified LEA proteins in multiple spots on 2D-gels belonging to two families; one with a sequence homology to group 3 LEA protein in *C. elegans* and the others with similarities to AtEm1 in plants [11].

As *A. franciscana* is a model organism for the study of adaptation mechanisms to extreme environmental conditions, in a quest to identify proteins that enable the dehydration tolerance of the encysted embryos, an analysis of heat-stable proteins was performed. Interestingly, a unique protein with an apparent mass of 21 kDa (p21) was prominent among the heat-stable proteins. Since to date *Artemia's* genome has not been sequenced, we utilized *de novo* mass spectrometry to identify a peptide fragment corresponding to p21. Furthermore, using MALDI-TOF we generated an amino acid sequence for the peptide that enabled the design of DNA oligonucleotides, which in turn

allowed the amplification of this gene from an *Artemia* embryonic cDNA library. This is the first report describing identification of a group 1 LEA protein from an animal species.

Materials and Methods

Dr. A. H. Warner and members of his research group provided an SDS-PAGE gel containing heat-stable protein p21 purified from *A. franciscana* encysted embryos. Details related to the purification protocol, MS-identification (described in detail below), PCR based amplification and DNA sequencing; as well as characterization of group 1 LEA protein (p21) were published by Sharon *et al.* [12].

De novo protein sequencing by MALDI-PSD

Following electrophoretic separation by SDS-PAGE and visualization by coomassie blue, the band corresponding to p21 was excised. Prior to in-gel proteolytic digestion, the gel piece was destained (50% acetonitrile, 50% 50 mM ammonium bicarbonate) for 1 hr at 37° C, dehydrated in 100% acetonitrile and excess liquid was removed in a speed vac. Thereafter, the gel piece was rehydrated in digestion buffer (50 mM ammonium bicarbonate) and digested with 200 ng of trypsin (Promega) for 12 hr at 37° C. Peptides were extracted (60% acetonitrile, 0.1% trifluoroacetic acid) and desalted using micro-C_{18} Zip Tips (Millipore). A portion of extracted peptides was subjected to chemically assisted fragmentation (CAF) modification in order to enhance the production of y-ions using the CAF Sequencing Kit according to manufacturer's instructions (Amersham Biosciences). Modified peptides were mixed at a ratio of 1:1 with matrix

solution (10 mg/mL α-cyano-4-hydroxycinnamic acid in 60% acetonitrile, 0.1% trifluoroacetic acid) on the target plate. Peptide mass fingerprints were obtained by matrix-assisted laser desorption-ionization-time of flight (MALDI TOF) in a reflector mode on a MALDI-TOF DE-Pro instrument (Applied Biosystems) using an accelerating voltage of 20,000V. MS spectra were internally calibrated using trypsin autolysis peptide masses. MS spectra were processed by use of the Data Explorer software (Version 4.0, Applied Biosystems) with a S/N threshold of 30 to detect peaks. Selected parent ions were subjected to post-source decay analysis using 11 segments stitched together by the Data Explorer software (Version 4.0, Applied Biosystems). Post-source decay spectra were calibrated with angiotensin I ($[MH]^+$ 1296.69) fragment ion masses. Owing to the lack of genomic sequences available for *A. franciscana*, *de novo* sequencing interpretation of the CAF-post source decay spectra was performed manually. A fragment mass tolerance of 0.3 Da was used to assign amino acid sequences. Peptide sequences were then searched against the NCBI non-redundant database using the NCBI BLAST program (version 2.2.14, www.ncbi.nlm.nih.gov/BLAST).

De novo protein sequencing by MALDI-MS/MS

In order to increase sequence coverage, an "in-gel" digestion of p21 was performed with either trypsin (Promega) (as described above) or 1 μg of endoproteinase GluC (Roche) for 12 hr at 25° C. Upon extraction, peptides were analysed on a 4800 MALDI TOF/TOF Analyzer (Applied Biosystems) operated in the positive ion reflectron mode. The MS data were internally calibrated using autolysis peaks of trypsin or GluC. Precursor ions with an S/N ratio of 30 were selected for tandem MS (MS/MS). All

MS/MS spectra were calibrated using MS/MS fragment masses from angiotensin I ($[MH]^+$ 1296.69). MS and MS/MS data were processed using the Data Explorer software (Version 4.5, Applied Biosystems). *De novo* sequencing was performed manually using a mass error of 0.2 Da between fragment ion masses to assign amino acid sequences. Once the DNA sequencing experiments were complete, the deduced protein sequence (entered into NCBI database under accession number EF656614) was used to corroborate MS data.

Results

Enrichment of a unique 21 kDa heat-stable cytoplasmic protein from *A. franciscana* dormant encysted embryos was initially observed after separation on SDS-PAGE. In order to identify this protein and obtain sequence information for cloning purposes, a MS-based approach was employed. The peptide fingerprint generated from trypsin digest, as well as the PSD-spectra acquired for a few peptides did not yield any useful matches when analysed against the publicly available databases (including NCBI ESTs). As a result of this and since the complete genome sequence of *A. franciscana* is not available, *de novo* sequencing was necessary in order to assign identity to the 21 kDa protein. Derivatization of the tryptic peptides by the CAF reagent (chemically assisted fragmentation) greatly simplified the interpretation of MS/MS spectra, as only y-ion series were detected, thus enabling amino acid assignment [13, 14]. The amino acid sequence of the CAF-modified peptide at $[MH^+]$ 1711 m/z was deduced as the following sequence: AE(K/Q)(I/L)GHEGYVEMGR (Figure 1A). A search of the NCBI database using the BLAST algorithm indicated that this 14 amino acid peptide showed significant

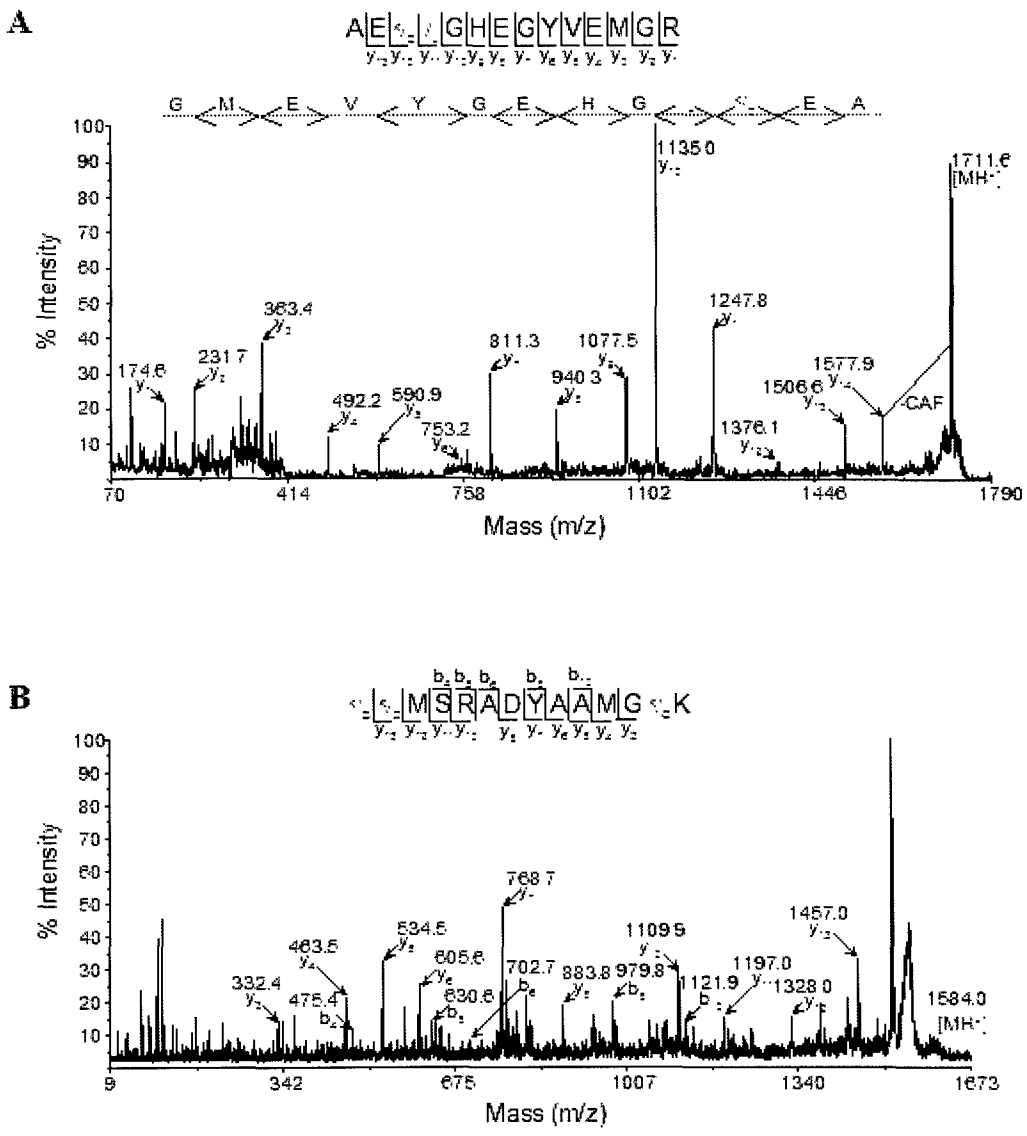


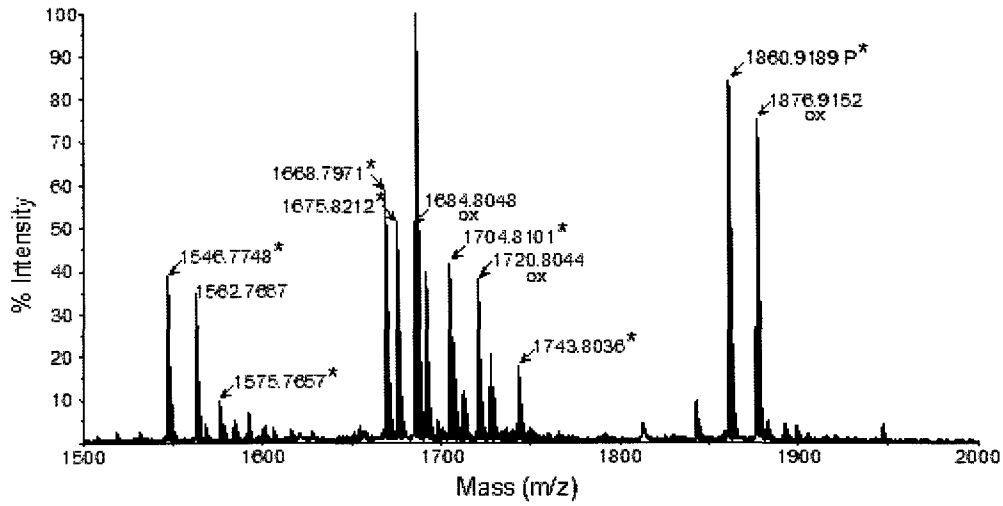
Figure 1. Analysis of the 21 kDa protein by matrix-assisted laser desorption-ionization (MALDI)-tandem mass spectrometry and *de novo* sequencing

(**A**) MALDI-post-source decay spectrum of chemically assisted fragmentation-modified tryptic peptide at m/z 1711.9. The y ion fragment ion series in annotated along with the loss of sulfonate group. (**B**) Tandem mass spectrometry spectrum of a tryptic peptide at m/z 1584.0 generated by a MALDI TOF/TOF instrument. The detected y and b ion fragment masses are annotated. For the deduced sequences, small font indicated the inability to differentiate between glutamine and lysine, or isoleucine and leucine.

similarity to the LEA family of proteins. Furthermore, this sequence was 71% identical to the group 1 LEA protein consensus sequence, thus far only described in few non-plant species [7, 8]. It is worth mentioning that the group 1 LEA protein sequence motif GGQTRREQLGEEGYSQMGRK is commonly present in several copies, ranging from up to four in plants and up to eight in other organisms [8]. The sequence of the CAF-modified peptide was used to design the original set of PCR primers for amplification of this gene from an embryonic cDNA library. The presence of repetitive sequence provided challenges during cloning efforts, as well as complicated assignments of unmodified MS/MS spectra. Fortunately, we detected a peptide ([MH^+] 1584.0 m/z) which was not a part of the repeating motif and were thus able to manually assign the amino acid sequence to (K/Q)(K/Q)MSRADYAAMG(K/Q)K (Figure 1B). It is noteworthy to mention, that this peptide was only detected in the unmodified trypsin digested sample, therefore the acquired spectra contained both, b- and y-ion fragment series, thereby further increasing the complexity of this assignment. The amino acid sequence of this peptide was used to design oligonucleotide primers to obtain the entire gene encoding the 21 kDa protein, including the C-terminal end.

Upon successful completion of the PCR-based amplification of the DNA sequence for group 1 LEA protein, the deduced sequence was used to obtain higher sequence coverage from the generated tryptic peptides (Figure 2). Overall, six peptides were subjected to MS/MS analysis, while four ([MH^+] 1546.8 m/z, 1668.8 m/z, 1675.8 m/z and 1704.8 m/z) of them corresponded to the originally identified CAF-modified peptide with variability in one to two amino acids. Therefore, these peptides represent the characteristic repeats found in group 1 LEA proteins. The fifth peptide ([MH^+] 1860.9 m/z) was an arginine miscleavage of one of the four peptides ([MH^+] 1704.8 m/z) and the

A



B

m/z	z	error [Da]	peptide
1546.7748	1	(+0.0066)	AE$^{K}/_{Q}{}^{I}/_{L}$GHEGYV$^{K}/_{Q}$MGK
1668.7971	1	(+0.0437)	AE$^{K}/_{Q}{}^{I}/_{L}$GTEGY$^{K}/_{Q}$EMG$^{K}/_{Q}$K
1675.8212	1	(+0.0468)	AE$^{K}/_{Q}{}^{I}/_{L}$GHEGYVEMG$^{K}/_{Q}$K
1704.8101	1	(+0.0455)	AE$^{K}/_{Q}{}^{I}/_{L}$GHEGY$^{K}/_{Q}$EMG$^{K}/_{Q}$K
1743.8036	1	(-1.0069)	$^{K}/_{Q}{}^{K}/_{Q}{}^{K}/_{Q}$mSRADYAAmG$^{K}/_{Q}$K
1860.9189	1	(+0.0526)	RAE$^{K}/_{Q}{}^{I}/_{L}$GHEGY$^{K}/_{Q}$EMG$^{K}/_{Q}$K

C

1 MSRADYAAMGQKAEQLGHEGYVEMGRKQQMSRADYAAMGQK

41 AEQLGHEGYQEMGQK

81 AEQLGHEGYVEMGQK

121 AEQLGTEGYQEMGQK

161 KQQMSRADYAAMGQK

AEQLGHEGYVEMGR

AEQLGTEGYQEMGQK

AEQLGHEGYQEMGQK

AEQLGHEGYVKMGK

Figure 2. Analysis of trysin generated peptides from the 21 kDa protein by matrix-assisted laser desorption-ionization (MALDI)-tandem mass spectrometry

(**A**) MALDI-TOF/TOF spectrum of the 21 kDa protein digested with trypsin. Peaks indicated with asterisks were subjected to MS/MS analysis. Peaks indicated with ox, represent methionine oxidation products. The peak at 1860, denoted with a P, represents a peptide containing polymorphic mutation. (**B**) Tabular presentation of peaks assigned for group 1 LEA protein sequence as determined by DNA sequencing from PCR-based cDNA library screen (for corresponding MS/MS spectra, see Figures 2D-2H) (EF6566124). (**C**) Protein sequence of group 1 LEA protein where the detected and sequenced peptides from the tryptic digest are underlined, while in boxes are identical peptide sequences found to be presented at multiple positions. Amino acids in grey colour correspond to the part of the protein from which peptides have not been detected or sequenced. Overall sequence coverage from the tryptic digest is 64.8%.

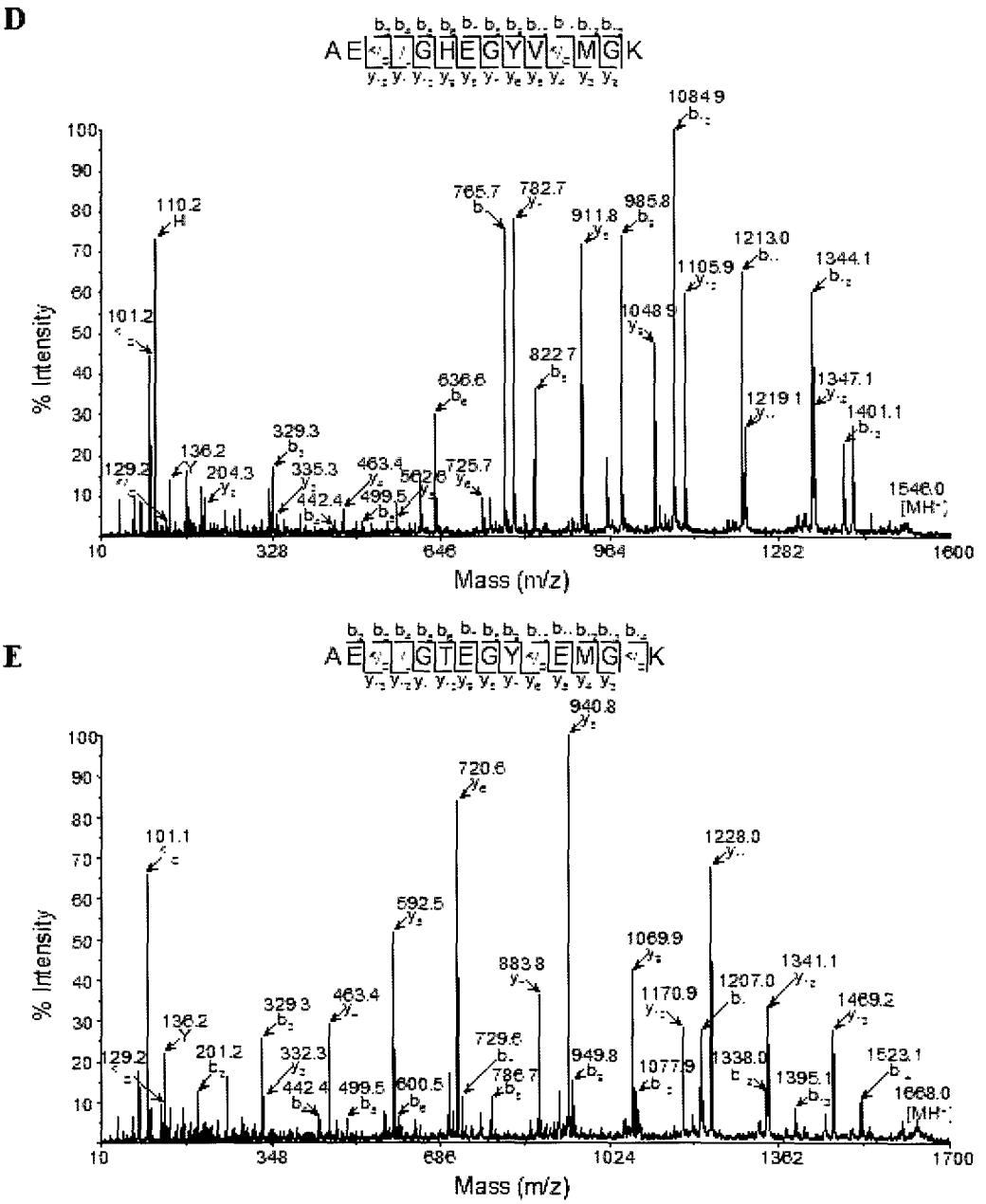


Figure 2. continued
MS/MS annotated spectra of tryptic peptides at 1546.0 m/z (**D**) and 1668.0 m/z (**E**).

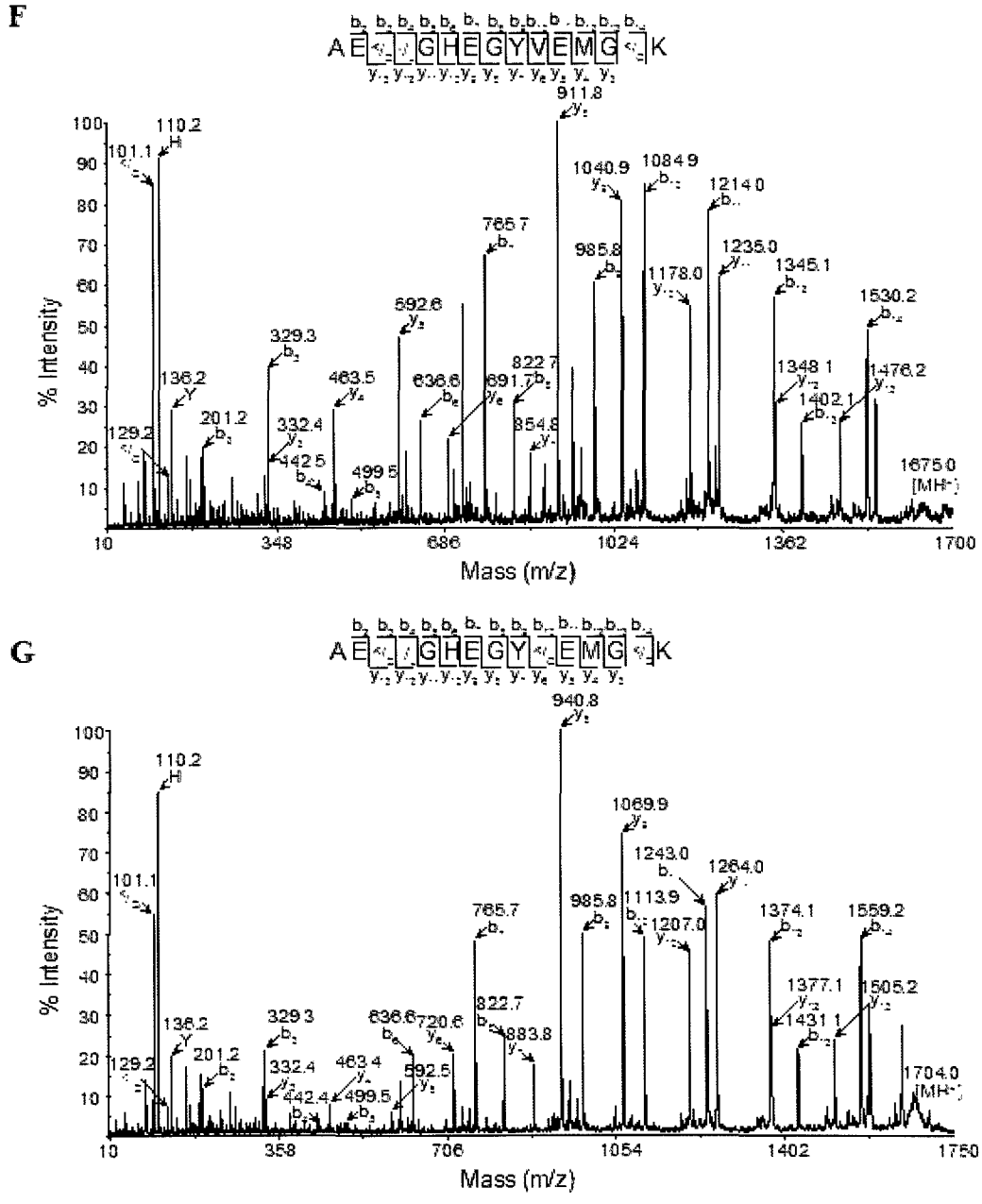


Figure 2. continued
MS/MS annotated spectra of tryptic peptides at 1675.0 m/z (**F**) and 1704.0 m/z (**G**).

H

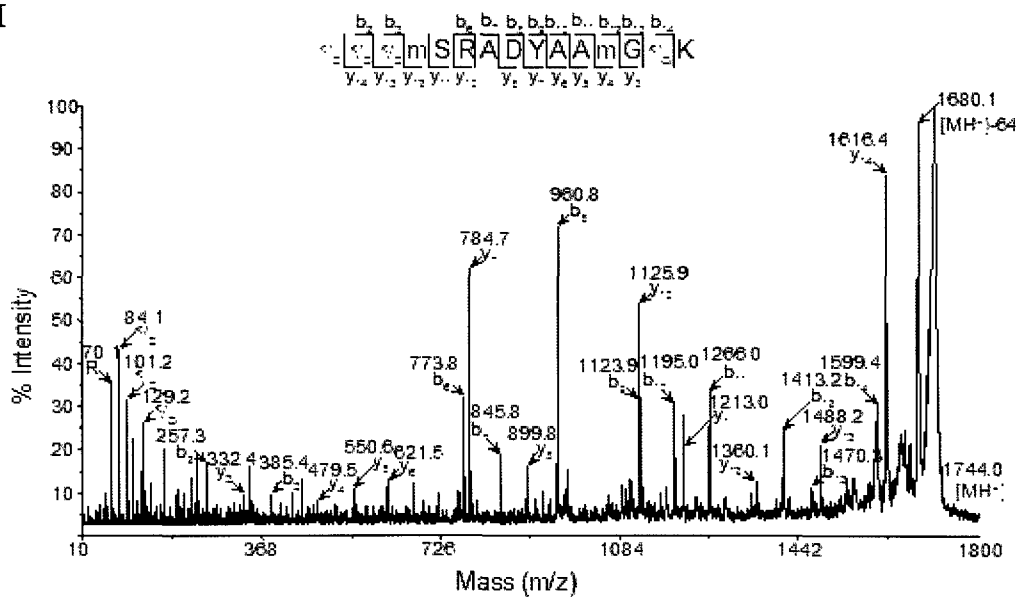


Figure 2. continued
MS/MS annotated spectra of tryptic peptide at 1744.0 m/z (**H**).

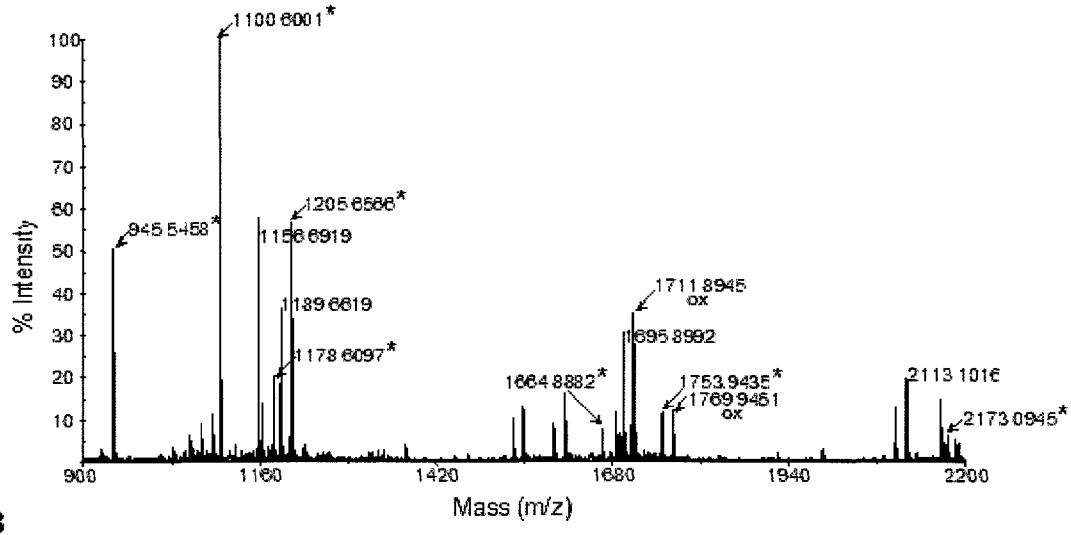
sixth peptide ([MH^+] 1743.8 m/z) was a lysine miscleavage and oxidized methionine version of the previously described unique sequence ([MH^+] 1584.0 m/z). Not surprisingly, sequences of two peptides identified by MS/MS ([MH^+] 1668.8 m/z and 1704.8 m/z) were present within the group 1 LEA protein in duplicates. The sequence coverage from tryptic digest is 64.8%.

In order to obtain more sequence information pertinent to group 1 LEA protein, Glu C generated peptide fragments were analyzed in similar fashion (Figure 3). Overall, seven peptides were subjected to MS/MS analysis, five were unique fragments representing the group 1 LEA repeats ([MH^+] 945.5 m/z, 1100.6 m/z, 1205.7 m/z, 1664.9 m/z and 2173.1 m/z), while one peptide ([MH^+] 1753.9 m/z) was the result of a GluC miscleavage ([MH^+] 1205.7 m/z). Similar to trypsin generated peptides, one GluC peptide ([MH^+] 1178.6 m/z) was assigned to the group 1 LEA protein three times. The sequence coverage from GluC digest is 51.1%. Combined sequence coverage from both proteolytic digests is 92.3%. Overall, a novel protein from *A. franciscana* encysted embryos (p21) was identified as a group 1 LEA protein by *de novo* sequencing. DNA sequence corresponding to this protein was corroborated by MS/MS data with high sequence coverage.

Discussion

As the studies on *A. franciscana* continue to shed light onto the molecular mechanism underlying tolerance to extreme environmental stress conditions, more protein families will be identified that facilitate survival of these embryos under virtually complete loss of cellular water for prolonged periods of time. Expression of LEA proteins

A



B

m/z	z	error [Da]	peptide
945.5458	1	(+0.0346)	$^{K}/_{Q}G^{K}/_{Q}{}^{I}/_{L}SR^{K}/_{Q}E$
1100.6001	1	(+0.0446)	$AG^{K}/_{Q}RGG^{K}/_{Q}ARAE$
1178.6097	1	(+0.0515)	$mG^{K}/_{Q}{}^{K}/_{Q}GG^{K}/_{Q}TRAE$
1205.6566	1	(+0.0511)	$mG^{K}/_{Q}{}^{K}/_{Q}GG^{K}/_{Q}{}^{K}/_{Q}RAE$
1664.8882	1	(+0.0670)	$AG^{K}/_{Q}RGG^{K}/_{Q}ARAE^{K}/_{Q}{}^{I}/_{L}GHE$
1753.9435	1	(+0.0673)	$MG^{K}/_{Q}{}^{K}/_{Q}GG^{K}/_{Q}{}^{K}/_{Q}RAE^{K}/_{Q}{}^{I}/_{L}GHE$
2173.0945	1	(+0.0378)	$GYVEMGR^{K}/_{Q}GG^{K}/_{Q}ARAE^{K}/_{Q}{}^{I}/_{L}GHE$

C

1 QGKLSRQ EAGQRGGQAR AEQLGHEGYV EMGRKGGQAR
41 AEQLGHE MGQKGGQKR
81 AEQLGHE MGQKGGQTR AE MGQKGGQTR
121 AE MGQKGGQTR AE
161

Figure 3. Analysis of GluC generated peptides from the 21 kDa protein by matrix-assisted laser desorption-ionization (MALDI)-tandem mass spectrometry

(**A**) MALDI-TOF/TOF spectrum of the 21 kDa protein digested with GluC. Peaks indicated with asterisks were subjected to MS/MS analysis. Peaks indicated with ox, represent methionine oxidation products. (**B**) Tabular presentation of peaks assigned for group 1 LEA protein sequence as determined by DNA sequencing from PCR-based cDNA library screen (for corresponding MS/MS spectra, see Figures 3D-3J) (EF6566124). (**C**) Protein sequence of group 1 LEA protein where the detected and sequenced peptides from the GluC digest are underlined, while in boxes are identical peptide sequences found to be presented at multiple positions. Amino acids in grey colour correspond to the part of the protein from which peptides have not been detected or sequenced. Overall sequence coverage from GluC digest is 51.1%.

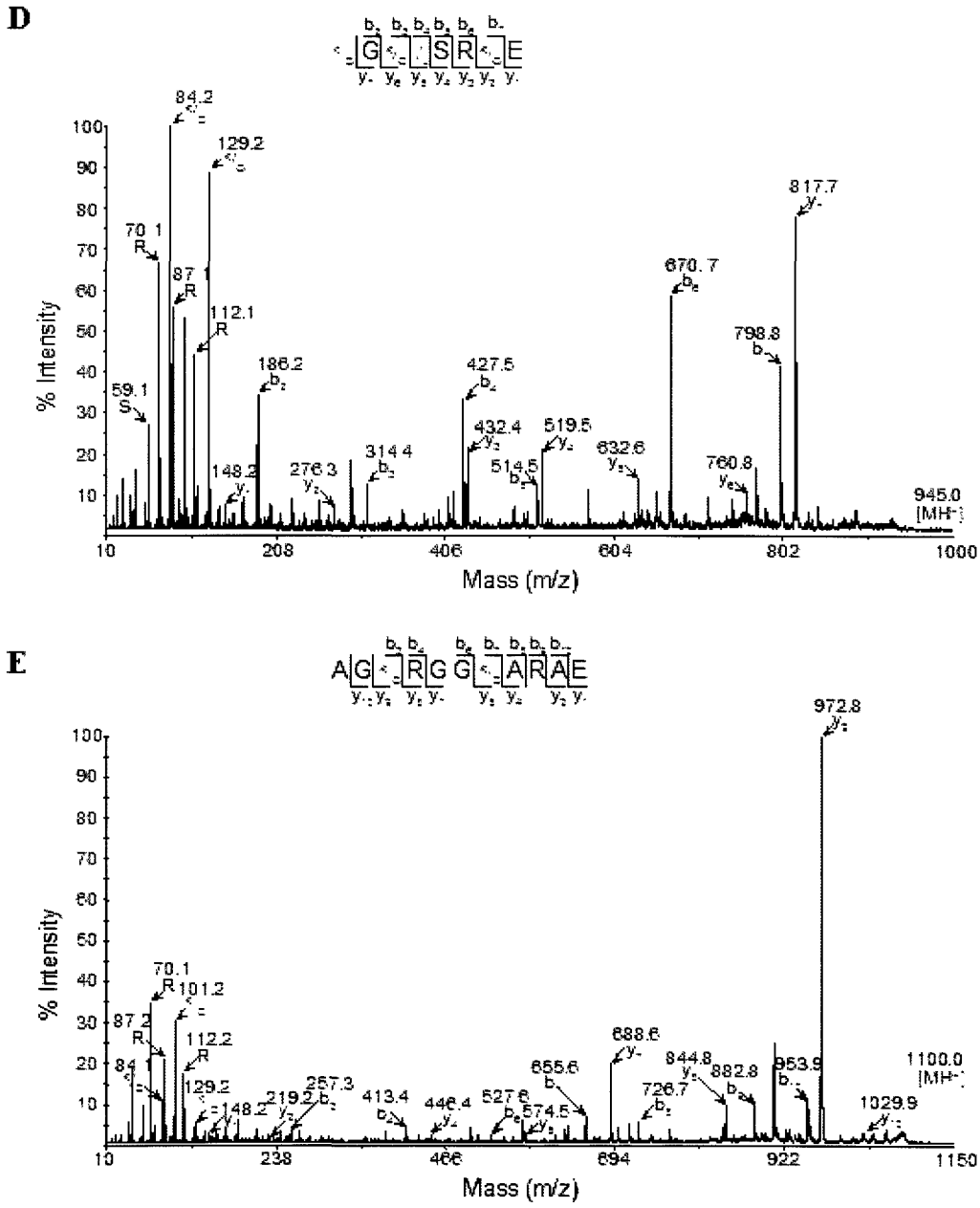


Figure 3. continued
MS/MS annotated spectra of GluC peptides at 945.0 m/z (**D**) and 1100.0 m/z (**E**).

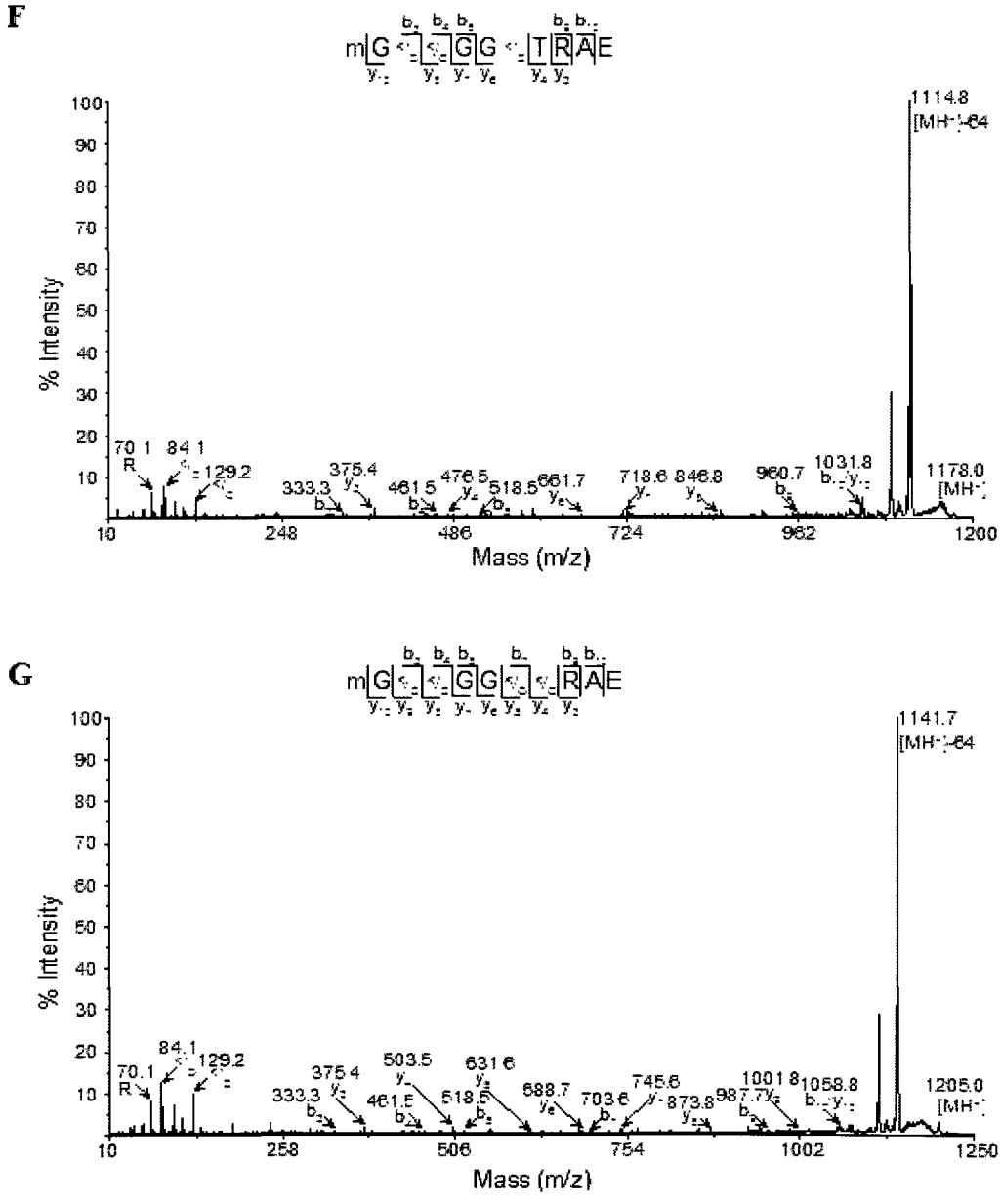


Figure 3. continued
MS/MS annotated spectra of GluC peptides at 1178.0 m/z (**F**) and 1205.0 m/z (**G**).

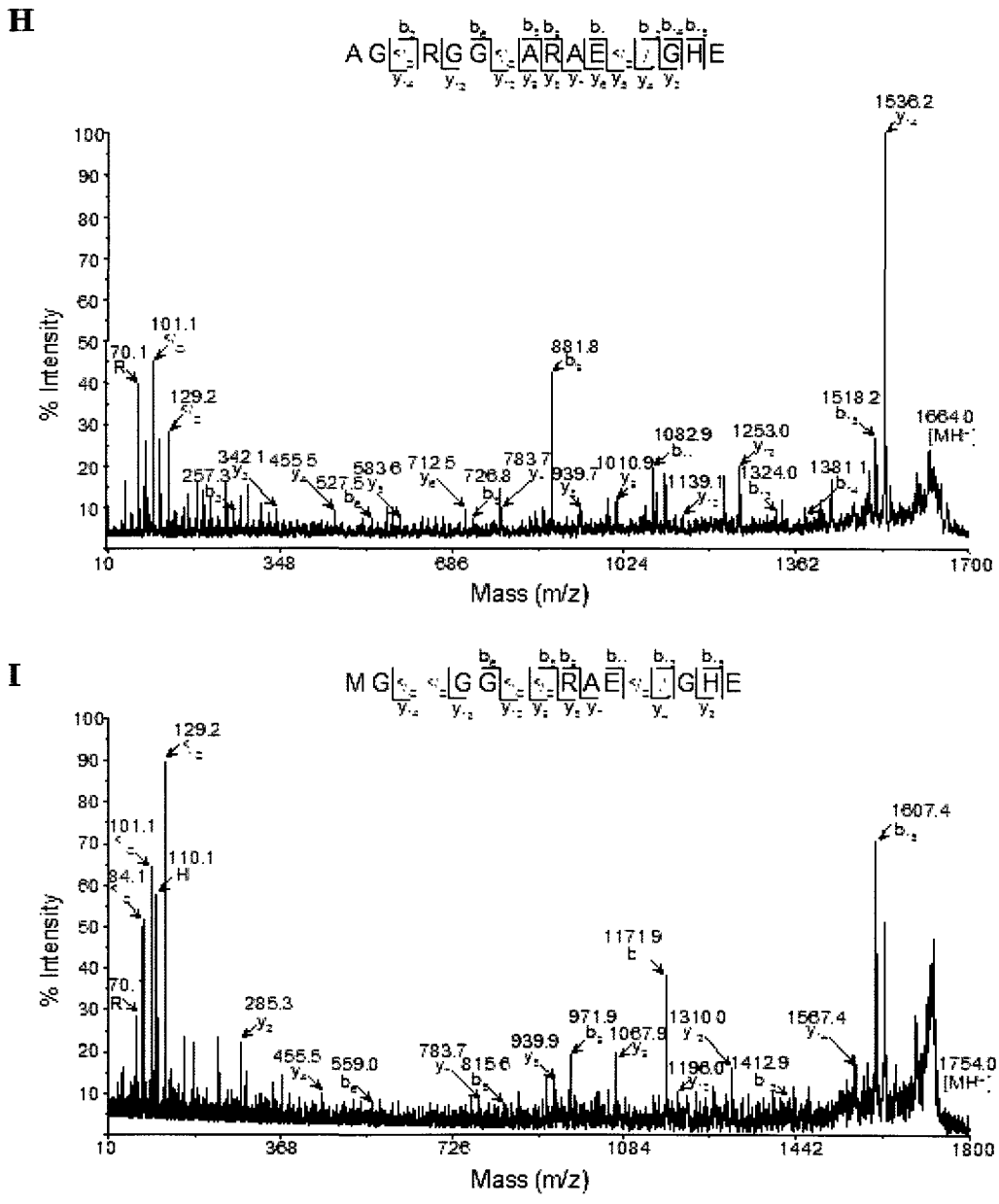


Figure 3. continued
MS/MS annotated spectra of GluC peptides at 1664.0 m/z (**H**) and 1754.0 m/z (**I**).

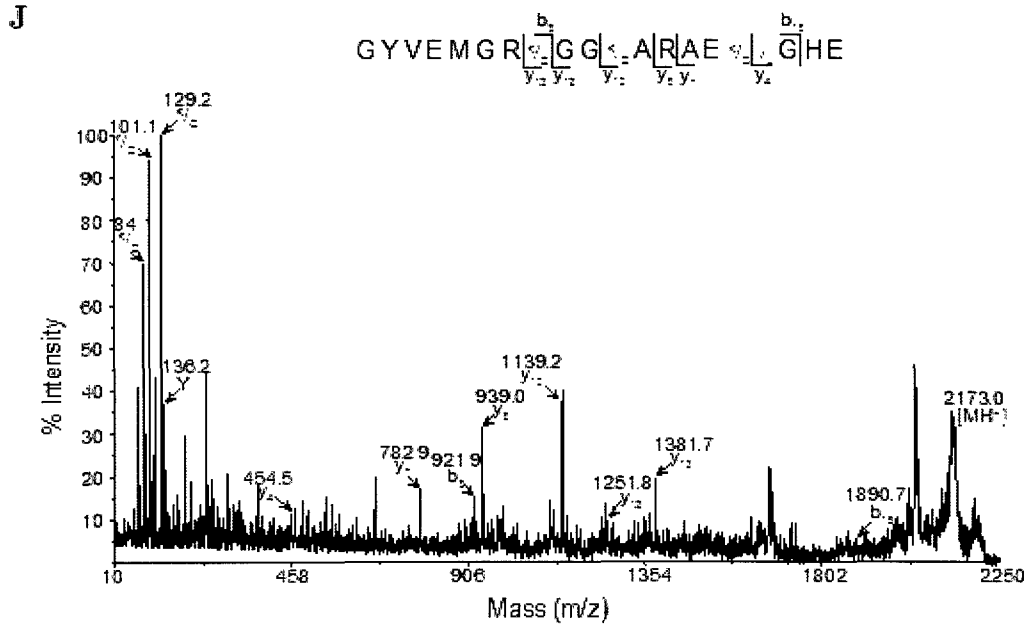


Figure 3. continued
MS/MS annotated spectra of GluC peptide at 2173.0 m/z (**J**).

often correlates with improved stress tolerance [7]. A heat-stable cytoplasmic p21 protein was identified as a group 1 LEA protein. Following identification, further analysis of this protein demonstrated a potential role for group 1 LEA proteins in desiccation tolerance of encysted embryos due to its ability to prevent drying-induced protein aggregation [12].

LEA family members are predicted to be generally small intrinsically unstructured proteins [7, 8]. The *A. franciscana* group 1 LEA protein is composed of 182 amino acids, from these 23% are glycine residues, which is in accordance with the group 1 LEA protein characteristic high content of glycine residues (approximately 18%) [8]. In addition, the sequence analysis revealed a large proportion of charged residues, contributing to the high hydrophilic nature of this protein. Furthermore, the sequence of the *A. franciscana* homologue lacks cysteines, a general characteristics of group 1 LEA proteins. Further supporting evidence that the identified protein is a group 1 LEA member is based on the regular spacing of glycine and glutamic acid residues within a period of twenty amino acids [7]. Sequence analysis of *A. franciscana* group 1 LEA protein is in agreement with this requirement.

The *A. franciscana* group 1 LEA protein contains eight tandem repeats of the 20 amino acid motif characteristic for group 1 LEA proteins. Overall, the repeats are very similar to each other. The first six repeats are 80% identical. The first and second and fourth and fifth repeat sequences are 90% identical, while 95% identity is observed between second and third and fifth and sixth repeat sequences. The tandem repeats become more degenerate towards the carboxy terminus. Repeats six and seven are 75% identical, while repeat seven and eight are only 35% identical. Similarly, a group 3 LEA protein, AfrLEA1, from *A. franciscana* displays divergence in nine characteristic repeats towards the carboxy terminus [10].

The presence of the tandem motifs described above greatly increased the complexity of the MS-spectra annotation. Many ionized peptides corresponded to the motif sequences with alterations in one or two amino acids, this caused by the variability among the repeats. Additionally, as the original heat stable protein sample was obtained from a population of encysted *A. franciscana* embryos, few polymorphisms were detected at the DNA levels as well as by the MS-peptide analysis. Most of the *de novo* MS/MS spectra were annotated manually, however once the PCR-based amplification of this gene from embryonic cDNA library was completed, the translated DNA sequence allowed for comparison between observed peptides and their corresponding amino acids with *in silico* generated fragments. *In silico* digest of the group 1 LEA protein was performed with MS-Digest software from Prospector (available at prospector.ucsf.edu). Other LEA proteins have been identified in a proteomic analysis on postdiapaused embryonic development in *A. franciscana* and were annotated based on matches within an in-house generated database containing 30,000 ESTs [11].

The *A. franciscana* mRNA for group 1 LEA is detected in diapaused embryos but not in swimming larvae [12]. While only encysted embryos are able to withstand desiccation for years, this implies a role of group 1 LEA protein in dessication tolerance. The identification of a unique 21 kDa protein by MS-based approach, in particular the *de novo* peptide sequence assignment to tandem motif, allowed cloning and characterization of the first group 1 LEA protein in animal species. *A. franciscana* group 1 LEA protein might be one of the still many missing pieces from the complex puzzle, which this organism has developed in order to protect their embryos from stressful environmental conditions.

Acknowledgments

This work was supported by Discovery Grants 2909 and 298468 to A.H.W and P.O.V., respectively, from the Natural Sciences and Engineering Research Council (NSERC) of Canada, and a postgraduate scholarship by NSERC to A.K.

We thank Dr. Philip Andrews and the Michigan Proteome Consortium, University of Michigan, for the use of their 4800 MALDI TOF/TOF instrument.

References

1. Clegg, J.S., *Protein stability in Artemia embryos during prolonged anoxia.* Biol Bull, 2007. **212**(1): p. 74-81.

2. Clegg, J.S., S.A. Jackson, and A.H. Warner, *Extensive intracellular translocations of a major protein accompany anoxia in embryos of Artemia franciscana.* Exp Cell Res, 1994. **212**(1): p. 77-83.

3. Liang, P. and T.H. MacRae, *The synthesis of a small heat shock/alpha-crystallin protein in Artemia and its relationship to stress tolerance during development.* Dev Biol, 1999. **207**(2): p. 445-56.

4. Warner, A.H., et al., *Artemin is an RNA-binding protein with high thermal stability and potential RNA chaperone activity.* Arch Biochem Biophys, 2004. **424**(2): p. 189-200.

5. Goyal, K., L.J. Walton, and A. Tunnacliffe, *LEA proteins prevent protein aggregation due to water stress.* Biochem J, 2005. **388**(Pt 1): p. 151-7.

6. Goyal, K., et al., *Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode.* FEBS Lett, 2005. **579**(19): p. 4093-8.

7. Tunnacliffe, A. and M.J. Wise, *The continuing conundrum of the LEA proteins.* Naturwissenschaften, 2007. **94**(10): p. 791-812.

8. Battaglia, M., et al., *The enigmatic LEA proteins and other hydrophilins.* Plant Physiol, 2008. **148**(1): p. 6-24.

9. Gal, T.Z., I. Glazer, and H. Koltai, *An LEA group 3 family member is involved in survival of C. elegans during exposure to stress.* FEBS Lett, 2004. **577**(1-2): p. 21-6.

10. Hand, S.C., et al., *Life without water: expression of plant LEA genes by an anhydrobiotic arthropod.* J Exp Zool A Ecol Genet Physiol, 2007. **307**(1): p. 62-6.

11. Wang, W., et al., *A proteomic study on postdiapaused embryonic development of brine shrimp (Artemia franciscana).* Proteomics, 2007. **7**(19): p. 3580-91.

12. Sharon, M.A., et al., *Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp Artemia franciscana.* Biochem Cell Biol, 2009. **87**(2): p. 415-30.

13. Keough, T., M.P. Lacey, and R.S. Youngquist, *Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using*

postsource decay matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom, 2000. **14**(24): p. 2348-56.

14. Keough, T., R.S. Youngquist, and M.P. Lacey, *A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry.* Proc Natl Acad Sci U S A, 1999. **96**(13): p. 7131-6.

CHAPTER 3

Identification of PDI redox sensitive thiols by ICAT-based MS

This chapter incorporates the outcome of joint research undertaken in collaboration with Inga Sliskovic under the supervision of Dr. Bulent Mutus, Department of Chemistry and Biochemistry. The collaboration is covered in Chapter 3 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the authors, and the contribution of co-authors was primarily through the provision of recombinant purified protein disulfide isomerase, PDI. Inga prepared the PDI protein sample in a reduced and auto-oxidized form. Dr. Eric Simon and Dr. Philip Andrews at the Michigan Proteome Consortium, University of Michigan, provided knowledge and instrumentation for the peptide separation and MALDI TOF/TOF analysis. My contribution to this manuscript was application of the isotope-coded affinity tag (ICAT) technology and identification of the cysteine redox state on PDI using mass spectrometry (MS). I performed all experimental steps involved in the ICAT labeling protocols, followed with acquisition and interpretation of the MS spectra.

The results were published: Kozarova, A., Sliskovic, I., Mutus, B., Simon, E.S., Andrews, P.C., and Vacratsis, P.O. (2007). Identification of redox sensitive thiols of protein disulfide isomerase using isotope coded affinity technology and mass spectrometry. J Am Soc Mass Spectrom *18*, 260-269

Introduction

Protein disulfide isomerase (PDI), a member of the thioredoxin superfamily of redox proteins, is an enzyme responsible for disulfide bond formation and isomerization [1, 2], redox-dependent chaperone function [3, 4] and S-nitrothiols denitrosation activity [5]. Originally, PDI was identified in the lumen of the endoplasmatic reticulum [6, 7] and subsequently detected at additional locations, such as the cell surface and the cytosol [8]. In the case of secreted PDI, it has been demonstrated to control the redox state of certain cell-surface proteins [9]. Furthermore, PDI was identified as a subunit of prolyl 4-hydroxylase [10, 11] and microsomal triglyceride transfer protein [12, 13], executing different functions than in its monomeric state.

PDI comprises four domains, ***a***, ***b***, ***b'*** and ***a'*** [14] and an acidic region designated ***c*** [15-17] (Figure 1A). The two thioredoxin domains of PDI [18], ***a*** and ***a'***, each contain an active site with a consensus sequence *WCGHCK*, which are separated by two thioredoxin-like domains ***b*** and ***b'***. The multi-domain structure of PDI is required to catalyze thiol-disulfide exchange events and effect folding by inducing conformational changes in the substrate protein [19]. Recently, the crystal structure of yeast PDI was elucidated and the structure revealed thioredoxin domains arranged in a shape of a twisted "U" with the active sites facing each other across the "U" [20].

Each PDI molecule contains six cysteine residues. Four of them have been shown to be responsible for the redox activity and are present as two vicinal thiol groups within each active site. Multiple studies have investigated the relative redox/catalytic contributions of the two thioredoxin active sites, and have suggested that they do not have equivalent catalytic properties [16, 21]. Furthermore, the NMR structure of human PDI ***a***

domain identified the N-terminal cysteine residue, Cys-36, as accessible and exposed while the C-terminal cysteine residue, Cys-39, as buried and constrained [22]. In addition, the crystal structure of yeast PDI predicted that the two cysteines within the ***a*** domain are primarily in the oxidized state (modeled as a mixture of oxidized to reduced state at 0.8 to 0.2 respectively) and the cysteines within the ***a'*** domain are in a reduced state [20]. The function and regulation of the redox states of the other two cysteine residues present within the ***b'*** domain of human PDI, remains unclear.

The redox state of active site cysteines dictates the catalytic activity of PDI. Opposite redox states are required to facilitate either oxidation or isomerization reactions, *i.e.* the best oxidases having the least stable disulfides and the best reductases having the least stable dithiols [1, 21, 23]. Therefore the thiol/disulfide exchange within the active site is required for effective oxidation or disulfide formation on PDI substrates [23]. Furthermore, alterations of the redox state balance of PDI underlie different pathological conditions. For example, it has been observed, that upon activation/aggregation of resting platelets the amount of reduced cell surface PDI increased approximately three fold, thus implying that thiol groups are critical for platelet aggregation [24].

Common approaches to analyze thiol oxidation states include spectrophotometric and fluorescence-based assays [25], however these only provide information regarding the average number of modified cysteine residues, not their exact identity. Here, we present a method that uses isotope-coded affinity tag based method (ICAT) in conjunction with mass spectrometry (MS) [26, 27] to directly detect and quantitate dynamic changes in the thiol oxidation state of recombinant human PDI under reducing and auto-oxidizing conditions. In addition to the detection of the redox status of the active site cysteines, we will offer data indicating redox regulation of Cys-295 present within the ***b'*** domain. Also,

a novel oxidation modification on an invariable tryptophan residue, Trp-35, adjacent to the cysteine residues within the active site, will also be presented.

Materials and Methods

Purification of Protein Disulfide Isomerase

His-PDI was expressed and purified as described previously [5]. Protein quantification was performed using the Bradford assay [28] and protein purity was ascertained by SDS-PAGE.

PDI reduction and spontaneous auto-oxidation

Bacterially expressed human His-PDI (100 μg) was fully reduced with 5x molar excess of DTT for 2 hr at room temperature. The DTT was removed during gel filtration on Sephadex G-25 in 100mM phosphate buffer, pH 7. Immediately after separation, the reduced PDI was acetone precipitated and labelled with light ^{12}C-ICAT reagent (Applied Biosystems). For spontaneous auto-oxidation, PDI was incubated for 1 hr at room temperature following the reduction with DTT and gel filtration. Subsequently, this sample was also acetone precipitated and labelled with heavy ^{13}C-ICAT reagent (Applied Biosystems).

ICAT labelling of reduced and auto-oxidized PDI

PDI containing samples were prepared as described above and ICAT labelling and purification of labelled peptides was performed following the provided protocol from

Applied Biosystems. Briefly, reduced and auto-oxidized PDI was labelled with ^{12}C-ICAT and ^{13}C-ICAT reagents, respectively, for 2 hr at 37°C. After alkylation with the iodoacetamide-based ICAT reagent, the two samples were mixed together and subjected to digestion with Endoproteinase *Glu-C* for 12 hr at 25°C. Peptides were desalted using a cation exchange cartridge and biotin labelled peptides were purified using an avidin affinity cartridge provided by Applied Biosystems. Peptides were subsequently dried and resuspended in the cleavage reagent in order to release the labelled peptides from the biotin moiety by incubation for 2 hr at 37°C. Subsequently, peptides were dried and resuspended in 1% formic acid. All samples were desalted using micro-C_{18} ZipTips (Millipore) prior to MALDI MS analysis.

HPLC separation of ICAT labelled peptides

ICAT labelled peptides were separated on a LC Packing nanoflow HPLC system (nano-LC) (LC Packings) with a Zorbax 300 SB C_{18} column (75μm x 150mm, 3.5μm particles) (Agilent technologies) at a flow rate of 300 nl/min. These peptides were separated using multi-step linear gradient of 0 to 30% B in the first 10 min, followed by 30 to 45% B for 37 min, 45 to 60% B for 5 min and 100% B for 6 min (mobile phase A, 3% acetonitrile, 0.1% trifluoroacetic acid; mobile phase B, 90% acetonitrile, 0.1% trifluoroacetic acid). Column effluent was mixed with MALDI matrix (2mg/ml α-cyano-4-hydroxycinnamic acid, 10mM ammonium dihydrogen phosphate in 50:49:1 water/isopropyl alcohol/acetic acid) through a mixing tee and spotted automatically with a Probot (LC Packings) onto a 192-well MALDI target plate. The matrix was delivered to the mixing tee at a flow rate of 1.2 μL/min.

Mass spectrometry analysis

ICAT labelled peptides were analyzed by MALDI-TOF MS (DE-Pro Voyager, Applied Biosystems) and MALDI-TOF/TOF (4700 Proteomics Analyzer, Applied Biosystems/MDX Sciex). For the MALDI-TOF MS analysis, the purified labelled peptide sample (1:1 matrix solution α-cyano-4-hydroxycinnamic acid) was spotted onto a MALDI target plate and analyzed in reflector mode. For the MALDI-TOF/TOF analysis MS and MS/MS spectra were acquired in positive ion reflector mode. MS survey scans were acquired first for each MALDI well and the ions of interest were selected for MS/MS analysis.

Results and Discussion

ICAT labelling of PDI

An ICAT-labelling approach in combination with mass spectrometry was developed for the purpose of quantitating changes in the redox state of human His-PDI under various conditions, since only free thiols will react with the iodoacetamide-based ICAT reagent. Bacterially expressed human PDI was purified to near homogeneity by nickel chromatography (Figure 1B). To assess if PDI can be labelled with ICAT reagent, following reduction with excess reducing agent, the free thiol groups on cysteine residues (cysteine-SH) were incubated with acid-cleavable light ^{12}C-ICAT reagent. The ICAT-labelled protein was digested with *Glu-C* in order to generate unique peptides corresponding to each active site. The [MH^+] 2147 m/z ion represents the peptide

containing the active site within the ***a*** domain, FYAPWC_{36}GHC_{39}KALAPE, containing two ICAT labels (molecular mass 227 Da per one ICAT-tag) (Figure 2A). This indicates that both cysteines within the ***a*** domain were in the reduced state and thus susceptible to labelling with the ICAT alkylating agent (confirmed by analysis of MS/MS fragment ion series, as described below). Importantly, we did not detect a peak at [MH^+] 1924 m/z, which would correspond to only one ICAT-labelled cysteine residue. Similarly to the ***a*** domain active site, a peptide fragment representative of the ***a'*** domain active site at [MH^+] 2916 m/z was modified with two ICAT labels. However, further analysis was complicated with the inability of reproducibly detecting this peptide. Thus, this implies that under these conditions, cysteine residues within the ***a'*** domain peptide FYAPWC_{380}GHC_{383}KQLAPIWDKLGE, exist predominantly in their oxidized form and are therefore less susceptible to ICAT modifications.

In addition, peptide fragments corresponding to the two remaining cysteine residues of the PDI within the ***b'*** domain were also detected (Table 1). Numerous peptides were detected encompassing the ICAT-labelled cysteine at position 295, specifically EC_{295}PAVRLITLE ([MH^+] 1470 m/z) and two peptides corresponding to additional mis-cleavages ([MH^+] 1729 m/z and [MH^+] 2579 m/z). The [MH^+] ion detected at 1178 m/z corresponded to the other cysteine within the ***b'*** domain, Cys-326, within the FC_{326}HRFLE peptide.

Mass spectrometric studies to analyze PDI have been performed in the past. An alternative to the ICAT-based labelling method is alkylation of the cysteine residues with N-ethylamaleimide (NEM) [29]. In parallel to the ICAT labelling described herein, we also modified fully reduced PDI with NEM. Fragments corresponding to active sites modified with one or two NEM adducts (125 Da per NEM label) were detected (data not

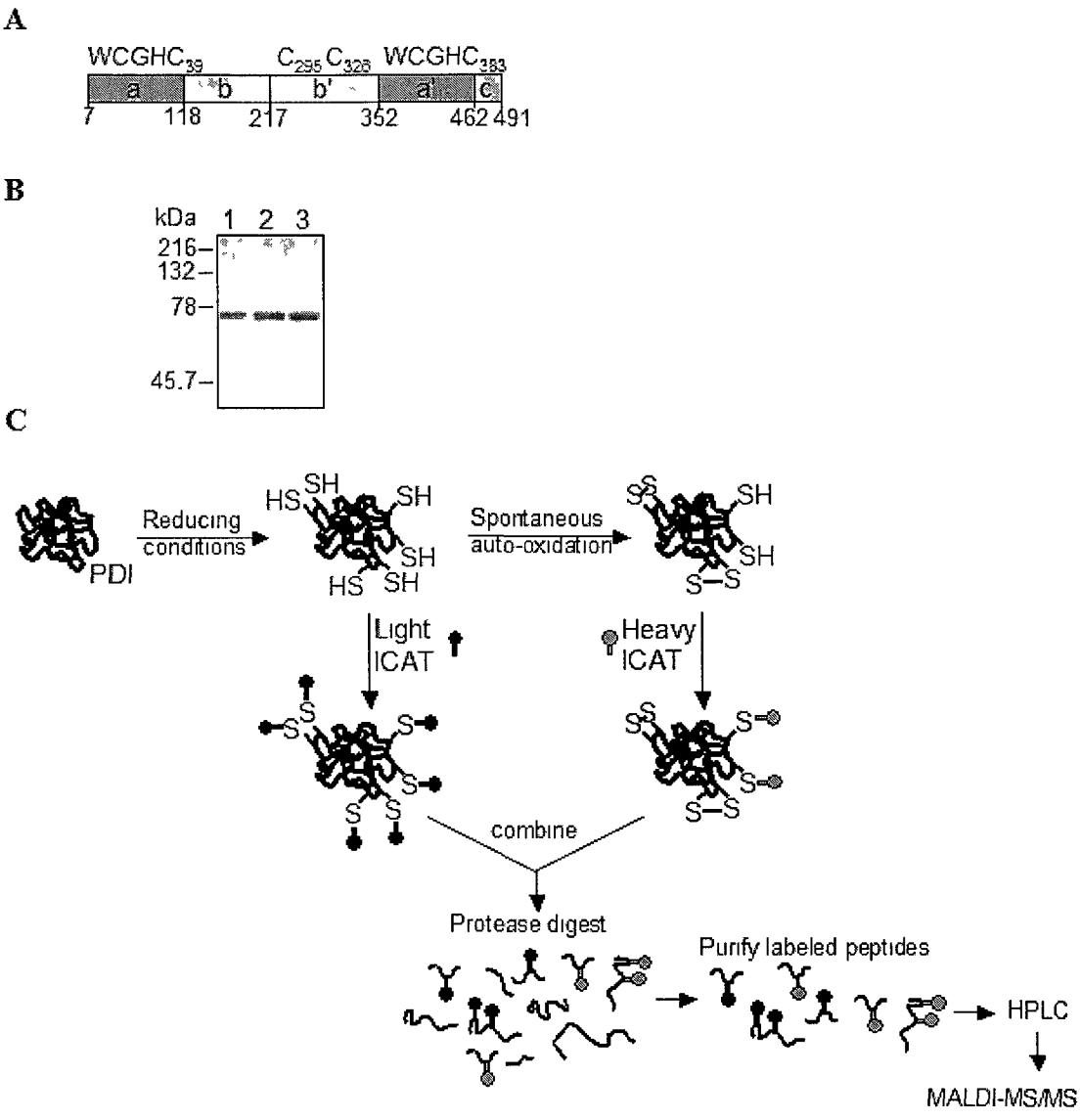


Figure 1. Schematics of ICAT strategy

(**A**) A schematic model of domain architecture of human PDI. Domains ***a***, ***b***, ***b'*** and ***a'*** share a common thioredoxin fold, but only the ***a*** and ***a'*** domain have the *WCXXC* active site. Position of all six cysteine residues within human PDI is highlighted. (**B**) Bacterially expressed, purified human His-PDI was separated using SDS-PAGE (10%) and visualized by commassie blue stain. *Lane 1* PDI control purchased from Sigma, *lane 2* reduced His-PDI (treated with 5x excess DTT), *lane 3* auto-oxidized His-PDI (following reduction, PDI was incubated in phosphate buffer for 1 hr at room temperature). (**C**) Schematic presentation of ICAT labelling protocol. PDI was reduced and immediately labelled with light ^{12}C-ICAT reagent or incubated at room temperature to allow auto-oxidation and subsequently labelled with heavy ^{13}C-ICAT reagent and processed as described in "*Materials and Methods*".

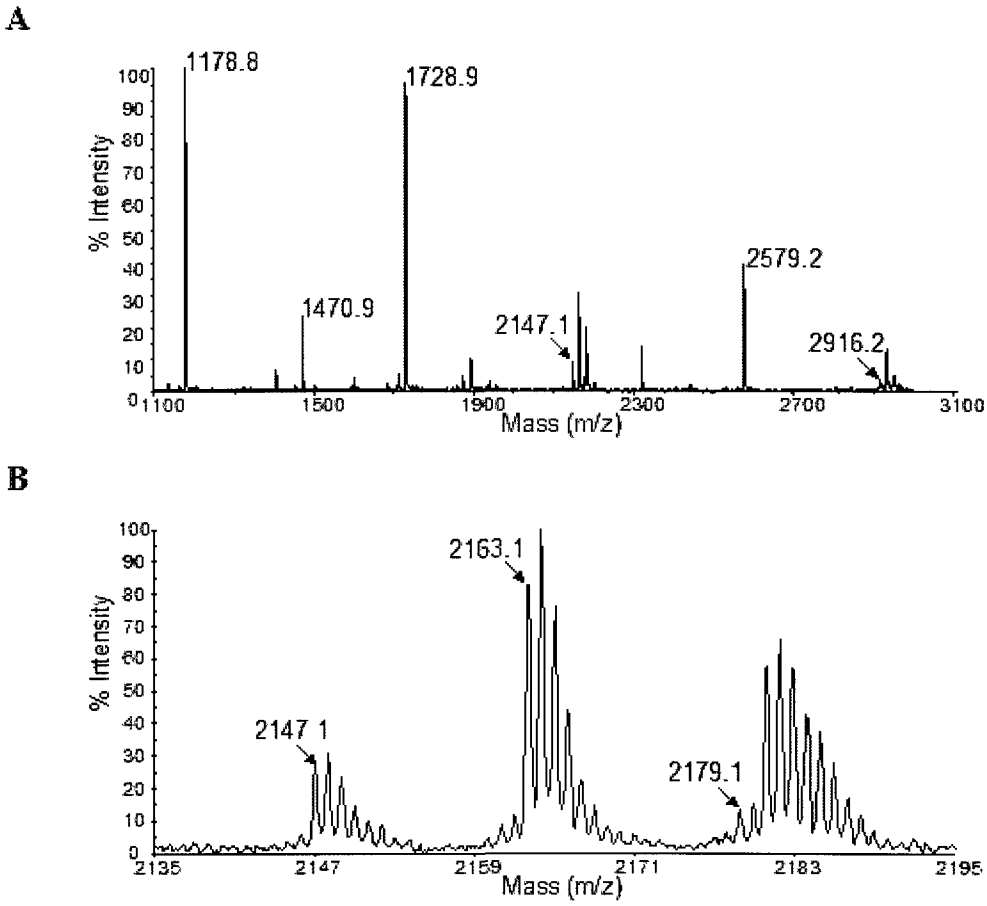


Figure 2. Reduced PDI labelled with light ^{12}C-ICAT reagent

(**A**) Full scan MALDI-TOF spectrum of ICAT-labelled *Glu-C* digested peptides from PDI. The [MH$^+$] 2147 m/z fragment corresponds to the ***a*** domain active site with both Cys residues ICAT labelled. The [MH$^+$] 2916 m/z fragment corresponds to the ***a'*** domain active site with both Cys residues ICAT labelled. The additional peptide fragments are presented in tabulated form, including the theoretical mass prior to Cys modification and after modification with corresponding detected ions, *i.e.* the fragment at [MH$^+$] 1178 m/z corresponds to the Cys-326 present in the ***b'*** domain while fragments at [MH$^+$] 1470 m/z, [MH$^+$] 1728 m/z and [MH$^+$] 2579 m/z represent various mis-cleavages of the sequence containing Cys-295. (**B**) Enlargement from A (range 2135-2195 m/z) showing the ions corresponding to addition of +16 Da and +32 Da on the ***a*** domain active site peptide.

#	Start	End	Cys	Sequence	Mass (monoisotopic) (m/z)	^{12}C-ICAT-modified (m/z) calculated	^{12}C-ICAT-modified (m/z) observed
1	325	331	326	FCHRFLE	951.5	1178.6	1178.8
2	294	304	295	ECPAVRLITLE	1243.6	1470.7	1470.9
3	294	306	295	ECPAVRLITLEEE	1501.8	1728.9	1729.0
4	31	45	36,39	FYAPWCGHCKALAPE	1692.7	2147.0	2147.1
5	287	306	295	FFGLKKEECPAVRLITLEEE	2351.2	2578.3	2579.3
6	375	395	380,383	FYAPWCGHCKQLAPIWDKLGE	2462.2	2916.4	2916.2

Table 1. Recombinant human PDI ICAT-labelled peptides

shown). In contrast, the ICAT-modified peptides reproducibly produced only peptides labelled on both cysteine residues. Another cysteine alkylating reagent, the episulfonium ion derived from S-(2-chloroethyl)glutathione (CEG) has been identified to modify only one cysteine residue within each active site of rat PDI, the N-terminal residues Cys-36 and Cys-380 [30]. In addition, cysteine residues of PDI were indicated to be modified by the thiol reactive lipid aldehyde 4-hydroxynonenal (4-HNE) in rat livers fed a combination of high-fat and ethanol diet [31], although the modified cysteines were not resolved. Overall, these different thiol reactivities may be due to incomplete labelling caused by various degrees of hydrophobicity and hydrophilicity of the modifying reagent, resulting in limited access to buried and solvent inaccessible regions within PDI. Our results indicate that the structural nature of the ICAT reagent is capable of reacting with all of the cysteine residues of PDI when they are present in their free thiol form.

Further analysis of the peptide corresponding to the ***a*** domain active site indicated two sequential peaks (each 16 Da apart) in addition to the peak analogous to the monoisotopic mass of [MH^+] 2147 m/z (Figure 2B). These modified light-ICAT labelled peptides would likely complicate the analysis when light and heavy ICAT-modified samples were to be compared. Specifically, the second generation ICAT reagents, ^{12}C-(light, addition of 227 Da per one ICAT-tag) and ^{13}C-(heavy, addition of 236 Da per one ICAT-tag), have been designed to differ by 9 Da per cysteine residue for quantitation purposes [27]. This would result in an addition of 18 Da when two ICAT labelled cysteines are present within the active site in the ***a*** domain. Therefore due to isotopic interference, the discrimination between the +16 Da (light ICAT-label and modified peptide ([MH^+] 2163 m/z)) and the + 18 Da (heavy ICAT-label and unmodified peptide ([MH^+] 2165 m/z)) would be challenging.

In order to resolve this issue, ICAT-labelled peptides were fractionated using reverse-phase chromatography (Figure 3). The column elution profile indicated that all of the modified peptides were significantly resolved from each other (Figure 3a), and in expected order. The peptide corresponding to the active site within the ***a*** domain labelled with 2 light ICAT-tags and modified by +32 Da eluted first (Figure 3b), followed by the peptide modified by +16 Da (Figure 3c), whereas the most hydrophobic, non-modified peptide eluted last (Figure 3c).

Tryptophan oxidation adjacent to the active site cysteine

MS/MS analysis identified the peptide at [MH^+] 2147 m/z to correspond to the ***a*** domain active site fragment, FYAPWC_{36}GHC_{39}KALAPE. This peptide contains two ICAT labels, one per cysteine residue, as determined from the y_6-y_{10} series, the y_{12} fragment and the b_8,b_{10}-b_{14} series (Figure 4A). Further analysis of the fragment ion series indicated that the modified residue on the peptide at 2163 m/z and the peptide at 2179 m/z was Trp-35. Specifically, the addition of +16 Da ([MH^+] 2163 m/z) was identified as tryptophan monooxidation (hydroxytryptophan) (Figure 4B) and the addition of +32 Da ([MH^+] 2179 m/z) was due to the tryptophan dioxidized form (*N*-formylkynurenine) (Figure 4C).

This is an intriguing modification since Trp-35 is located within the ***a*** domain active site immediately adjacent to the active site cysteines. Moreover the modified forms of the peptide were present in greater yield than the unmodified form (Figure 2). It is possible, that this modification was the result of MS sample preparation. However, it has been argued that tryptophan oxidation is not a considered mass spectrometry encountered artifact, like methionine oxidation [32, 33] and guanidination [34]. Additionally, doubly

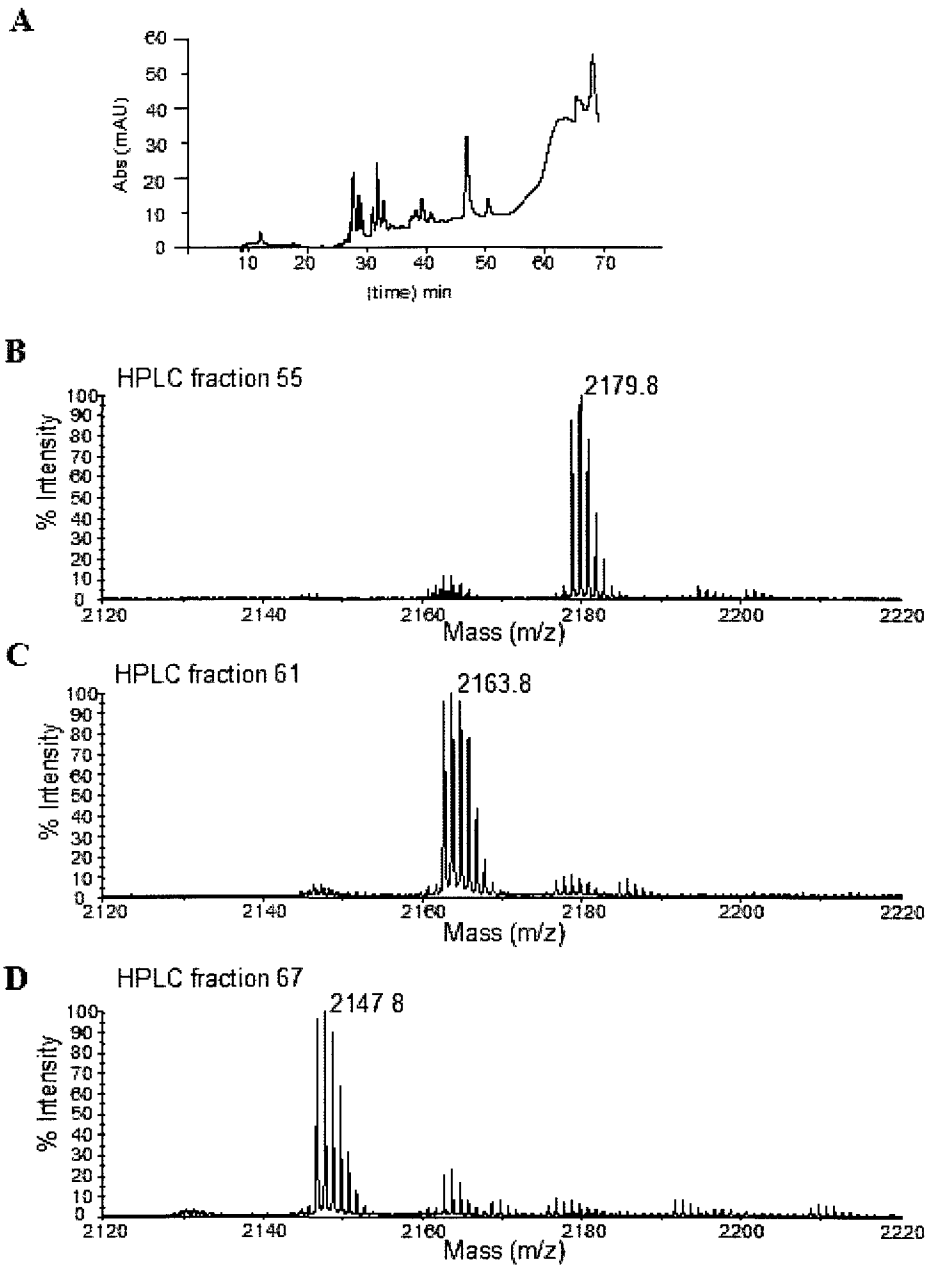


Figure 3. Separation of the modified ^{12}C-ICAT-labelled peptides
(**A**) Chromatogram showing separation of the *Glu-C* digested, ICAT labelled peptides on a reverse-phase nano-LC column. (**B**) MALDI-TOF analysis of peptides separated by nano-LC. The doubly modified peptide (+32 Da) eluted first ([MH^+] 2179 m/z), followed by singly modified peptide (+16 Da) ([MH^+] 2163 m/z) (**C**), and the unmodified peptide ([MH^+] 2147 m/z) (**D**).

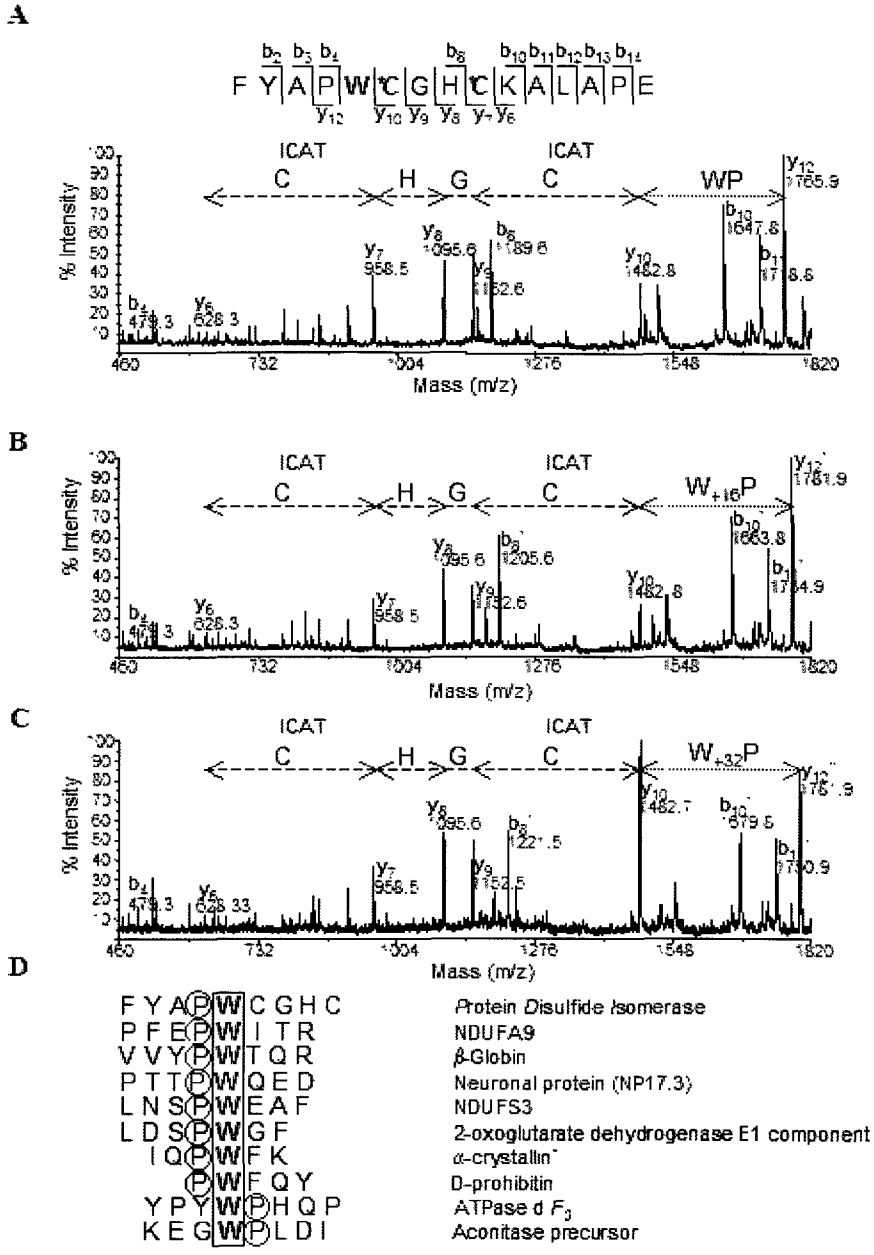


Figure 4. MALDI-TOF MS/MS analysis of the *a* domain active site labelled with light ^{12}C-ICAT reagent displaying tryptophan oxidation

(**A**) MS/MS spectrum of the unmodified peptide ([MH$^+$] 2147 m/z) showing the amino acid sequence and position of ICAT modifications. (**B**) The +16 Da modified peptide ([MH$^+$] 2163 m/z) was determined to contain a hydroxytryptophan. (**C**) The +32 Da modified peptide ([MH$^+$] 2179 m/z) was determined to contain a doubly oxidized tryptophan, N-formylkynurenine. Asterisks within the sequence indicates modification of cysteine residues with ICAT reagent,' indicates modification of the y- and b-ions by +16 Da (B) and '' indicates modification by + 32 Da (C). (**D**) Sequence alignment of peptides identified to contain doubly oxidized tryptophan; these peptides were identified in the cardiac mitochondrial proteome [33] and from bovine intact lens (*)[35].

oxidized tryptophan was identified in a subset of mitochondrial proteins predominantly associated with redox metabolism [33] and in the chaperone α-crystallin from bovine lens tissue exposed to Fenton chemistry [35]. Sequence analysis of the residue surrounding Trp-35 within PDI suggests similarity with these other proteins modified *in vivo* by tryptophan oxidation. In addition, the location of this tryptophan residue immediately adjacent to the active site cysteines suggests that tryptophan oxidation may play a role in regulating PDI catalysis.

Dynamic changes in the redox state of PDI active site

We have previously shown that PDI undergoes auto-oxidation in the absence of reducing agents [5]. In order to further characterize and definitively identify the cysteine residues altered during auto-oxidation, the ICAT strategy was used to quantify changes in the redox state of PDI. Specifically, reduced PDI was labelled with the light ^{12}C-ICAT reagent, while auto-oxidized PDI was labelled with the heavy ^{13}C-ICAT reagent (Figure 1C). These samples were subsequently combined and processed together. Upon fractionation of PDI peptides by nano-LC, the relative quantity of thiol oxidation was assessed by MS. For the purified ***a*** domain peptides, all modified forms displayed approximately a 5 to 1 ratio between reduced and auto-oxidized samples (Figure 5). The ratio of non-modified tryptophan containing peptides labelled with ICAT-light reagent (([MH^+] 2147 m/z) can be compared to non-modified tryptophan containing peptides labelled with ICAT-heavy reagent ([MH^+] 2165 m/z) (Figure 5C). The ratio of the light and heavy ICAT labelled hydroxytryptophan containing peptides ([MH^+] 2163 m/z and [MH^+] 2181 m/z) was consistent to the unmodified peptides (Figure 5B) as was the case

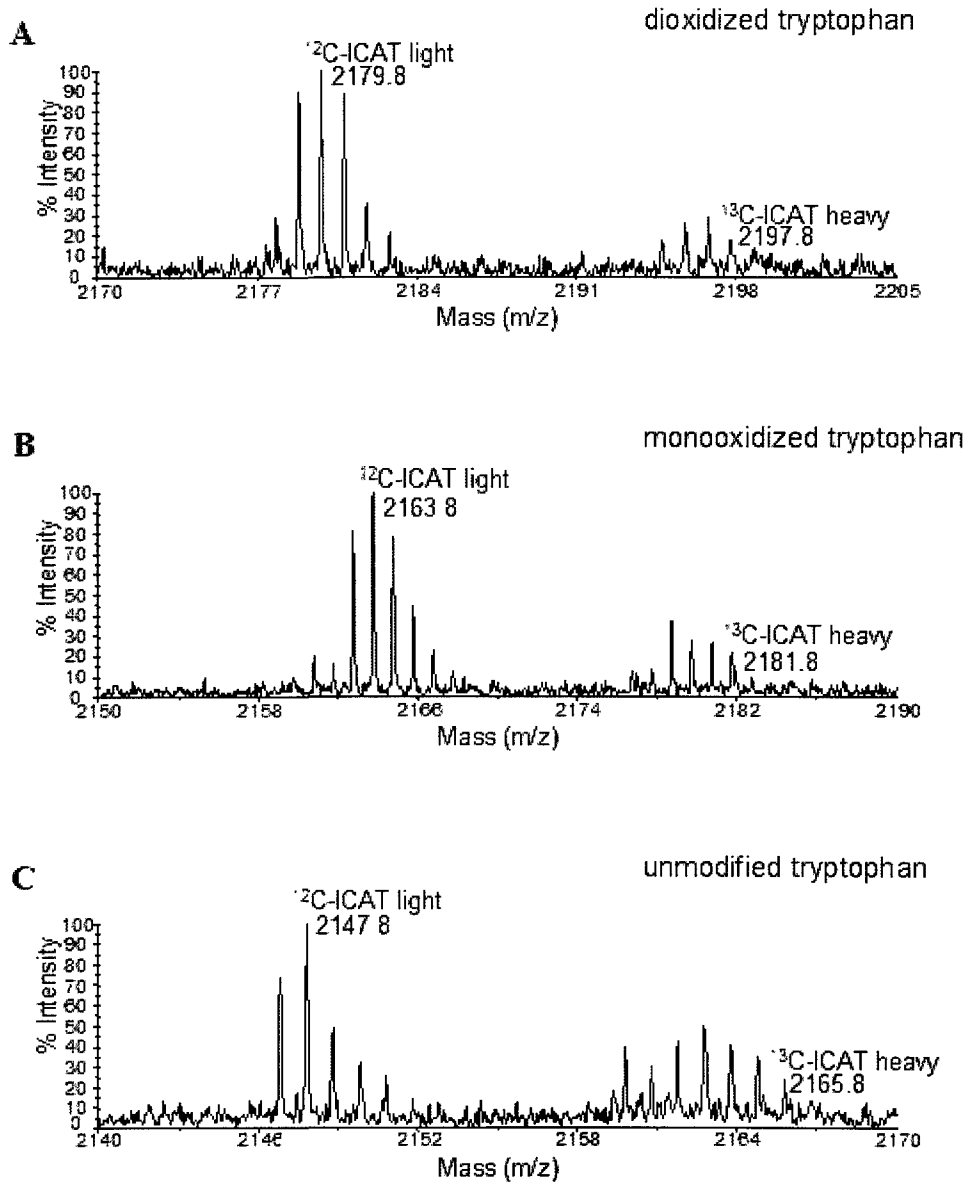


Figure 5. Quantitation of reduced (light ^{12}C-ICAT-labelled) and auto-oxidized (heavy ^{13}C-ICAT-labelled) PDI following reverse-phase nano-LC separation

(**A**) Peptides corresponding to the ***a*** domain active site containing the doubly oxidized Trp, were detected for reduced PDI (light ICAT labelled, [MH^+] 2179 m/z) and auto-oxidized PDI (heavy ICAT labelled, [MH^+] 2197 m/z) displaying approximately 5 fold decrease in susceptibility to labelling. Peptides corresponding to the ***a*** domain active site containing the monooxidized Trp ([MH^+] 2163 m/z and 2181 m/z) (**B**) and non-modified Trp ([MH^+] 2147 m/z and 2165 m/z) (**C**) were detected for reduced PDI and auto-oxidized PDI displaying similar auto-oxidation patterns.

of N-formylkynurenine containing peptides ([MH^+] 2179 m/z and [MH^+] 2197 m/z) (Figure 5A).

Redox-induced changes of b' domain cysteines

So far, the functional significance of the two cysteine residues within the ***b'*** domain of the human PDI has not been completely elucidated. Point-mutation studies indicate that these residues do not play any role in PDI catalytic activity or substrate interaction [19, 36-38]. We observed, that auto-oxidation does not have any effect on Cys-326, within the sequence FC_{326}HRFLE ([MH^+] 1178 m/z and [MH^+] 1187 m/z), and therefore serves as an internal control for the reduced and auto-oxidized samples (Figure 6).

Interestingly, auto-oxidation induced changes in the redox state of the other ***b'*** domain cysteine residue, Cys-295. Two differential cleavage peptides containing Cys-295 ([MH^+] 1470 m/z and [MH^+] 1479 m/z; [MH^+] 1728 m/z and [MH^+] 1737 m/z) demonstrated a dramatic loss of ICAT reactivity under auto-oxidized conditions (Figure 6). This result suggests, that in addition to changes within the ***a*** domain active site of human PDI, Cys-295 within the ***b'*** domain is also sensitive to redox regulation.

The ICAT-based approach used in this study indicated redox regulation of one of the two cysteine residues within the ***b'*** domain. Specifically, the ability to modify Cys-295 with ICAT-reagent upon auto-oxidation decreased about five fold. In contrast, Cys-326 remained susceptible to ICAT-labelling even after auto-oxidation, thus implying that this cysteine residue remains in thiol form. This is interesting since a study involving chemical modification of an isolated ***b'*** domain indicated that neither of these cysteine residues are likely to exchange between dithiol and disulfide states upon modulation of

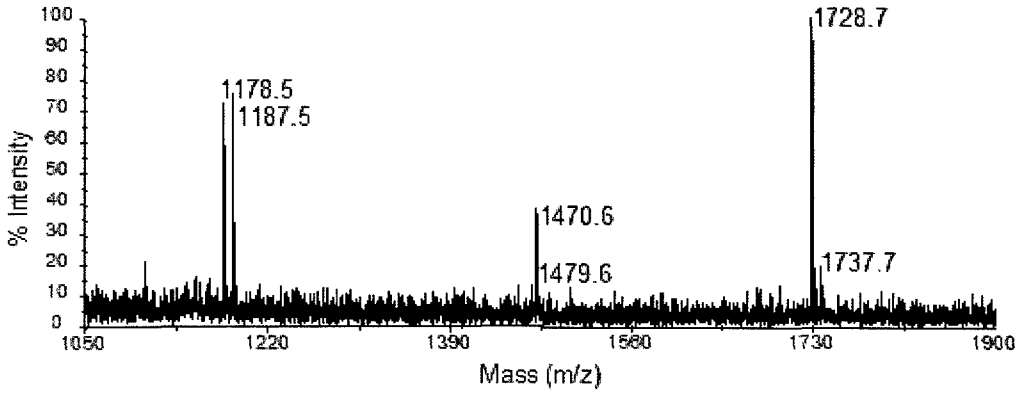


Figure 6. Redox regulation of Cys-295 within the *b'* domain
MALDI-TOF analysis of peptides containing ***b'*** domain Cys residues labelled with light and heavy ICAT. Similar levels of peptides containing Cys-326 from reduced (light ICAT) and auto-oxidized (heavy ICAT) samples were detected at [MH^+] 1178 m/z and [MH^+] 1187 m/z respectively. The Cys-295 containing peptides from two mis-cleavage products ([MH^+] 1470 m/z and [MH^+] 1728 m/z for light ICAT) show a five fold decrease in ICAT-labelling upon auto-oxidation ([MH^+] 1479 m/z and [MH^+] 1737 m/z).

redox potential [37]. In contrast, based on the crystal structure of yeast PDI, Cys-90 and Cys-97, were elucidated to form a disulfide bond [20]. These two non-active site cysteines are part of the ***a*** domain of yeast PDI (unlike the ***b'*** domain of human PDI) and are believed to destabilize the oxidated state of the ***a*** domain active site cysteines [39]. Our results using full length PDI implicated Cys-295 as an additional redox sensitive site, potentially capable of modulating PDI activities.

This study reports direct detection and conclusive identification of redox induced changes on the cysteine residues of recombinant human PDI in the ***a***, ***b'*** and ***a'*** domains using ICAT methodology combined with mass spectrometry. The ICAT data presented here agree with the predicted susceptibility to oxidation of the active site vicinal thiols within the ***a*** domain, in that both of these cysteine residues were reactive towards ICAT-reagent in the fully reduced enzyme. However, the ICAT-signal of these thiols decreased by five fold upon auto-oxidation. In contrast, the ***a'*** active site cysteine residues could not be reliably labelled with ICAT-reagent, and are thus probably present in an oxidized state. This is in accordance with the decrease of total cysteine content upon auto-oxidation as determined by reactivity with 5,5'-dithiobis(nitrobenzoic acid) [5]. In conclusion, PDI activity is redox dependent and regulated; therefore we developed a method to study direct dynamic changes in the redox state of PDI.

Acknowledgment

We thank John Hudson for critical reading of the manuscript. This project has been funded by a Grant-in-Aid from the Canadian Diabetes Association to B.M. and P.O.V. and a postgraduate scholarship by the Natural Sciences and Engineering Research Council (NSERC) to A.K.

References

1. Gilbert, H.F., *Protein disulfide isomerase and assisted protein folding.* J Biol Chem, 1997. 272(47): p. 29399-402.

2. Ellgaard, L. and L.W. Ruddock, *The human protein disulphide isomerase family: substrate interactions and functional properties.* EMBO Rep, 2005. 6(1): p. 28-32.

3. Tsai, B., et al., *Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin.* Cell, 2001. 104(6): p. 937-48.

4. Janiszewski, M., et al., *Regulation of NAD(P)H oxidase by associated protein disulfide isomerase in vascular smooth muscle cells.* J Biol Chem, 2005. 280(49): p. 40813-9.

5. Sliskovic, I., A. Raturi, and B. Mutus, *Characterization of the S-denitrosation activity of protein disulfide isomerase.* J Biol Chem, 2005. 280(10): p. 8733-41.

6. Munro, S. and H.R. Pelham, *A C-terminal signal prevents secretion of luminal ER proteins.* Cell, 1987. 48(5): p. 899-907.

7. Shin, B.K., et al., *Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function.* J Biol Chem, 2003. 278(9): p. 7607-16.

8. Turano, C., et al., *Proteins of the PDI family: unpredicted non-ER locations and functions.* J Cell Physiol, 2002. 193(2): p. 154-63.

9. Jiang, X.M., et al., *Redox control of exofacial protein thiols/disulfides by protein disulfide isomerase.* J Biol Chem, 1999. 274(4): p. 2416-23.

10. Pihlajaniemi, T., et al., *Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene.* Embo J, 1987. 6(3): p. 643-9.

11. Koivu, J., et al., *A single polypeptide acts both as the beta subunit of prolyl 4-hydroxylase and as a protein disulfide-isomerase.* J Biol Chem, 1987. 262(14): p. 6447-9.

12. Wetterau, J.R., et al., *Protein disulfide isomerase is a component of the microsomal triglyceride transfer protein complex.* J Biol Chem, 1990. 265(17): p. 9801-7.

13. Lamberg, A., et al., *The role of protein disulphide isomerase in the microsomal triacylglycerol transfer protein does not reside in its isomerase activity.* Biochem J, 1996. 315 (Pt 2): p. 533-6.

14. Edman, J.C., et al., *Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin.* Nature, 1985. 317(6034): p. 267-70.

15. Koivunen, P., et al., *The acidic C-terminal domain of protein disulfide isomerase is not critical for the enzyme subunit function or for the chaperone or disulfide isomerase activities of the polypeptide.* Embo J, 1999. 18(1): p. 65-74.

16. Darby, N.J., J. Kemmink, and T.E. Creighton, *Identifying and characterizing a structural domain of protein disulfide isomerase.* Biochemistry, 1996. 35(32): p. 10517-28.

17. Freedman, R.B., T.R. Hirst, and M.F. Tuite, *Protein disulphide isomerase: building bridges in protein folding.* Trends Biochem Sci, 1994. 19(8): p. 331-6.

18. Lundstrom, J. and A. Holmgren, *Protein disulfide-isomerase is a substrate for thioredoxin reductase and has thioredoxin-like activity.* J Biol Chem, 1990. 265(16): p. 9114-20.

19. Darby, N.J., E. Penka, and R. Vincentelli, *The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency.* J Mol Biol, 1998. 276(1): p. 239-47.

20. Tian, G., et al., *The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites.* Cell, 2006. 124(1): p. 61-73.

21. Lyles, M.M. and H.F. Gilbert, *Mutations in the thioredoxin sites of protein disulfide isomerase reveal functional nonequivalence of the N- and C-terminal domains.* J Biol Chem, 1994. 269(49): p. 30946-52.

22. Kemmink, J., et al., *Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy.* Biochemistry, 1996. 35(24): p. 7684-91.

23. Xiao, R., et al., *Catalysis of thiol/disulfide exchange. Glutaredoxin 1 and protein-disulfide isomerase use different mechanisms to enhance oxidase and reductase activities.* J Biol Chem, 2005. 280(22): p. 21099-106.

24. Burgess, J.K., et al., *Physical proximity and functional association of glycoprotein 1balpha and protein-disulfide isomerase on the platelet plasma membrane.* J Biol Chem, 2000. 275(13): p. 9758-66.

25. Wright, S.K. and R.E. Viola, *Evaluation of methods for the quantitation of cysteines in proteins.* Anal Biochem, 1998. 265(1): p. 8-14.

26. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. 17(10): p. 994-9.

27. Hansen, K.C., et al., *Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography.* Mol Cell Proteomics, 2003. 2(5): p. 299-314.

28. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. 72: p. 248-54.

29. Yen, T.Y., H. Yan, and B.A. Macher, *Characterizing closely spaced, complex disulfide bond patterns in peptides and proteins by liquid chromatography/electrospray ionization tandem mass spectrometry.* J Mass Spectrom, 2002. 37(1): p. 15-30.

30. Kaetzel, R.S., et al., *Alkylation of protein disulfide isomerase by the episulfonium ion derived from the glutathione conjugate of 1,2-dichloroethane and mass spectrometric characterization of the adducts.* Arch Biochem Biophys, 2004. 423(1): p. 136-47.

31. Carbone, D.L., et al., *Cysteine modification by lipid peroxidation products inhibits protein disulfide isomerase.* Chem Res Toxicol, 2005. 18(8): p. 1324-31.

32. Thiede, B., et al., *Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion.* Rapid Commun Mass Spectrom, 2000. 14(6): p. 496-502.

33. Taylor, S.W., et al., *Oxidative post-translational modification of tryptophan residues in cardiac mitochondrial proteins.* J Biol Chem, 2003. 278(22): p. 19587-90.

34. Karty, J.A., et al., *Artifacts and unassigned masses encountered in peptide mass mapping.* J Chromatogr B Analyt Technol Biomed Life Sci, 2002. 782(1-2): p. 363-83.

35. Finley, E.L., et al., *Identification of tryptophan oxidation products in bovine alpha-crystallin.* Protein Sci, 1998. 7(11): p. 2391-7.

36. Schwaller, M., B. Wilkinson, and H.F. Gilbert, *Reduction-reoxidation cycles contribute to catalysis of disulfide isomerization by protein-disulfide isomerase.* J Biol Chem, 2003. 278(9): p. 7154-9.

37. Pirneskoski, A., et al., *Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase.* J Biol Chem, 2004. 279(11): p. 10374-81.

38. Klappa, P., et al., *The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins.* Embo J, 1998. 17(4): p. 927-35.

39. Wilkinson, B., R. Xiao, and H.F. Gilbert, *A structural disulfide of yeast protein-disulfide isomerase destabilizes the active site disulfide of the N-terminal thioredoxin domain.* J Biol Chem, 2005. 280(12): p. 11483-7.

CHAPTER 4

Identification of hYVH1 phosphorylation sites by MS and their function

This chapter incorporates the outcome of joint research undertaken in collaboration with Dr. John W. Hudson, Department of Biological Sciences. The collaboration is covered in Chapter 4 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the authors, and the contribution of co-authors was through the provision of experimental help, analysis and data interpretation with flow cytometry based approach. My contribution to this manuscript in preparation was identification of phosphorylation sites on Flag-hYVH1 expressed and immunoprecipitated from mammalian cells; generation of phosphomutants, localization experiments, RNAi experiments and characterization of the phenotype, as well as sample processing for flow cytometry experiments.

Introduction

Protein phosphorylation (by protein kinases) and dephosphorylation (by protein phosphatases) is a universal regulatory mechanism controlling a vast array of signalling pathways and physiological processes [1-4]. Dual specificity phosphatases (DUSPs) encompass a group of heterogeneous and versatile enzymes that are able to remove phosphate moieties from phosphotyrosine and phosphoserine/threonine residues [5]. The best characterized role of DUSPs is the regulation of proline directed kinase families such as the mitogen activated protein kinases (MAPKs) and the cyclin dependent kinases (CDKs).

Atypical dual specificity phosphatases represent a DUSP subfamily and consist of generally smaller enzymes (only three members of this subgroup are bigger than 27 kDa), and lack the N-terminal Cdc25 homology 2 domain (which is typical for DUSPs) [5]. Therefore, it is postulated that most atypical DUSPs are not direct regulators of MAPKs or CDKs. Furthermore, certain atypical DUSPs contain unique domain architectures that likely dictate distinctive cellular functions; *e.g.* laforin contains a carbohydrate binding domain [6], DUSP 11 has arginine- and proline-rich regions [7] and DUSP12 encompasses a zinc binding domain [8, 9].

Human YVH1, (hYVH1, DUSP12) contains a conserved N-terminal catalytic domain and a novel C-terminal zinc-binding domain [9]. Thus it is not surprising, that upon clustering analysis of human atypical dual specificity phosphatases, hYVH1 appears on a separate branch of the dendogram [10]. Disruption of yeast DUSP12 (YVH1) causes a slow growth phenotype [11], which can be complemented by the human orthologue (hYVH1), or a mutant containing a catalytically inactivating version of hYVH1 (hYVH1-C115S). This ability to rescue the mutant phenotype is lost upon deletion of the C-

terminal zinc binding domain [9]. Additional phenotypes due to deletion of YVH1 in yeast include defects in glycogen metabolism and spore maturation [12]. A yeast two-hybrid screen identified YPH1 (yeast pescadillo homolog, Nop7) as a yeast YVH1 interacting partner [13]. Nop7 is required for multiple activities during 60S ribosomal synthesis [14]. Interestingly, YVH1 was also recently identified as a 60S trans-acting factor in a mass spectrometry (MS)-based analysis of pre-60S particles [15]. It was further shown that interplay between YVH1 and Mrt4 is required for the late maturation step of the 60S subunit [16, 17], thus highlighting the role of YVH1 in the ribosomal biogenesis pathway.

Human YVH1 was found by a MS-based approach to interact with Hsp70 [18]. Over-expression of hYVH1 repressed cell death induced by cellular insults, such as heat shock, hydrogen peroxide exposure and Fas receptor activation. This repression is further enhanced upon co-expression with Hsp70 [18]. Also, a role for hYVH1 in tumourigenesis is supported by a recent clinical study, in which hYVH1 was shown to be selectively and highly overexpressed in $CD34^+$ progenitor cells from chronic myelogenous leukemia patients [19]. However, hYVH1 remains a poorly understood phosphatase especially regarding how its activities are regulated in the cell.

Here we report the first identified phosphorylation sites within hYVH1 using a MS based approach. The sites were further characterized in terms of their effect on hYVH1 localization and DNA content of cells over-expressing the phosphomimetic and non-phosphorylatable mutants. In addition, the cell cycle phenotype observed due to down-regulation of hYVH1 is also described.

Materials and methods

DNA plasmids

Flag-hYVH1 and C115S were described previously [9]. Domain specific constructs, *i.e.* Flag-hYVH1$_{1\text{-}192}$ (ΔCT1), Flag-hYVH1$_{1\text{-}268}$ (ΔCT2), Flag-hYVH1$_{1\text{-}282}$ (ΔCT3), and Flag-hYVH1$_{193\text{-}340}$ (Zn-binding domain) were described previously [18].

The mutants S335A and S335E, were generated by replacement of hYVH1 amino acids sequence 331-340 flanked by *SspI* and *BamHI* sites with synthetic oligonucleotide duplexes containing the mutation, sequences as follow; for S335A oligo1 5'ATTGCCTG TTTTGGGAGCACAAACAGGAAAGATCTGAG 3', oligo2 5'GATCCTTCAGATCTT TCCTGTTTGTGCTCCCAAAACAGGCAAT 3'; for S335E oligo1 5'ATTGCCTGTTT TGGGAGAACAAACAGGAAAGATCTGAG 3' and oligo2 5'GATCCTCAGATCTTT CCTGTTTGTTCTCCCAAAACAGGCAAT 3'.

The S14A and S14E mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene). Primers are as follows, for S14A primer1 5'GATGGCTGCGAGCTCGCCAACCCCAGCG 3', primer2 5'CGCTGGGGTTGGCG AGCTCGCAGCCATC 3'; and for S14E primer1 5'GATGGCTGCGAGCTCGAGAAC CCCAGCGCCAG 3' and primer2 5'CTGGCGCTGGGGTTCTCGAGCTCGCAGCCAT C 3'. Multiple mutations were introduced in stepwise manner using the site-directed mutagenesis approach.

Flag-hYVH1 phosphomimetic TP mutants were generated by QuickChange site-directed mutagenesis kit (Stratagene) following the manufacturer protocol. Primers are as follows, for T252A primer1 5'GCCCACAAGAGAATGGCACCATCTTCCATG 3',

primer2 5'CATGGAAGATGGTGCCATTCTCTTGTGGGC 3'; and for T252E primer1 5'CCCACAAGAGAATGGAACCATCTTCCATGCTTACCAC 3', primer2 5' GTGGT AAGCATGGAAGATGGTTCCATTCTCTTGTGGG 3'.

All mutations were verified by DNA sequencing.

Tissue culture, transfections and treatments

HEK 293 (ATCC) cells were maintained in Dulbecco's modified eagle's medium nutrient mixture F-12 HAM (Sigma) supplemented with 10% fetal bovine serum (FBS, Sigma) and 2mM glutamine (Gibco); HeLa cells (ATCC) were grown in DMEM (Thermo Scientific) containing 10% FBS at 37°C and 5% CO_2. Transient transfections were performed either by Fugene HD (Roche) following manufacturer directions or by calcium phosphate method [20] for 24-48 hr. For treatment with phosphatase inhibitors, cells were exposed to calyculin A (Cal A; 100nM; LC Laboratories) for 30 minutes, okadaic acid (OA; 5μM; LC Laboratories) for 30 minutes, cyclosporin A (CsA; 10μM; LC Laboratories) for 24 hr or dimethylsulphoxide (solvent), prior to lysis.

For MAPK inhibitors, cells were treated with SB 203580 (10μM, LC Laboratories), SP600125 (40μM, LC Laboratories), U0126 (10μM, LC Laboratories) for 1 hr prior to lysis. For combination experiments, first cells were treated with MAPK inhibitors for 30 min prior to addition of Cal A, and then both inhibitors were incubated together for an extra 30 min.

Immunoprecipitation experiments

Transfected cells were lysed in 50mM Tris 7.5, 150mM NaCl, 1% Triton X-100, 0.1% SDS containing 100 $ug.ml^{-1}$ aprotinin, 1mM phenylmethylsulphonyl fluoride and 50mM sodium fluoride, and then centrifuged at 20,000g for 30 minutes at 4°C. The cleared cell lysates were incubated with pre-equilibrated anti-Flag M2 agarose (Sigma) for 3 hr, followed by three washes in wash buffer (50mM Tris 7.5, 150mM NaCl, 0.1% Triton X-100, 0.1% SDS). For dephosphorylation experiments, immunoprecipitated Flag-hYVH1 was treated with 30U calf intestinal alkaline phosphatase (CIP, Promega) for 30 minutes at 37°C. Proteins were resolved on 10% SDS-PAGE gels and stained first with Pro-Q Diamond phosphoprotein stain (Molecular Probes) using manufacturer provided protocol and visualized on Molecular Imager FX (Bio-Rad). Thereafter, gels were either silver stained [21] or coomassie brilliant blue G stained.

Preparation of phosphopeptides and MS analysis

Peptides were prepared by in-gel digestion of immunoprecipitated Flag-hYVH1 separated on SDS-PAGE gel [22]. Briefly, bands corresponding to Flag-hYVH1 were excised and incubated in destain solution (50% acetonitrile, ACN, HPLC-grade Burdick & Jackson, and 50% 50mM ammonium bicarbonate, Fisher) for 35 minutes at 37°C. Afterwards, gel pieces were dehydrated in 100% ACN and dried for 20 minutes in a speed vac. Protein were digested with either Endoproteinase GluC (Roche) or sequencing grade modified trypsin (Promega) following manufacturer directions. Peptides were collected by extraction with 60% ACN and 1% formic acid. Collected peptides were desalted on C_{18} ZipTip (Millipore) or C_{18} spin columns (Vivascience) following

manufacturers supplied protocol prior to MALDI-TOF analysis (Applied Biosystems Voyager, DE PRO instrument).

In order to enrich for low abundance phosphopeptides, desalted peptides were subjected to PHOS-select^{TM} Iron Affinity Gel (Sigma), using the manufacturer provided protocol. Briefly, peptides were incubated with pre-equilibrated bead slurry (in 250mM acetic acid, 30% ACN). Non phosphorylated peptides were removed (250mM acetic acid, 30% ACN) and phosphopeptides were eluted (400mM ammonium hydroxide). Samples were desalted as describe above. Peptides were mixed with matrix (10 mg/ml α-cyano-4-hydroxycinnamic acid in 60% ACN and 1% formic acid), and spotted onto MALDI-plate. MS instrument was calibrated with Sequazyme peptide mass standards kit (Applied Biosystems). The peptide finger-print was acquired in positive linear mode, with the accelerating voltage set to 25,000V, grid voltage set to 94%, guide wire to 0.05% and the extraction delay time to 100 nsec. Spectra were analyzed using the Data Explorer Software (Applied Biosystems). Experimental peptide masses were compared to theoretical values calculated by ProteinProspector MS-Digest Program (available at http://prospector.ucsf.edu).

siRNA experiments

Endogenous hYVH1 was down-regulated by transfection with Lipofectamine RNAiMAX (Invitrogen) of Stealth siRNA duplex oligonucleotides (CCUGGAAACUAU GCUUUACAUGGCA; Invitrogen) for 48 hr, following manufacturer provided protocol. As a negative control, cells were transfected with Stealth RNAi Negative universal control low GC duplex #3 (Invitrogen). Prior to immunoblot analysis, Bradford assay

(BioRad) was performed to determine the total amount of proteins in pre-cleared lysates. Proteins resolved on SDS-PAGE were transferred electrophoretically to PVDF membrane, which were probed with optimal dilutions of primary antibodies (endogenous anti-hYVH1 [9], anti-actin (Sigma)), followed by secondary antibody (anti-rabbit HRP, BioRad). Antibody conjugates were detected by SuperSignal west femto maximum sensitivity substrate (Thermo Scientific) according to manufacturer's directions.

Immunofluorescence analysis

HeLa cells grown on glass cover-slips were fixed in phosphate-buffered saline (PBS) containing 3.7% paraformaldehyde and 0.1% Triton X-100 for 12 minutes at room temperature. After washing with PBS, cells were permeabilized in PBS containing 0.1% Triton X-100 for 2 minutes at room temperature. Prior to incubation with antibodies, cells were blocked by 10% BSA for 30 minutes at 37°C. Cells were the incubated with anti-Flag M2 monoclonal antibody (Sigma) at 1:500 in Tris-buffered saline containing 0.1% Tween-20 (TBST) for 1 hr at room temperature, followed by washes in TBST. Primary antibody was visualized by incubation with fluorescein-conjugated anti-mouse (Vector Labaratories) at 1:500 for 30 minute at room temperature. DNA was counterstained with Hoechst 33342 (Molecular Probes) prior to mounting with SlowFade Antifade (Molecular Probes). Images were acquired on Zeiss Axioskop 2 (*mot plus*) microscope equipped with QImagin Qcam camera and analyzed with Northern Eclipse software.

Flow cytometry analysis

Adherent HEK 293/HeLa cells expressing Flag-hYVH1 and various mutants were trypsinized (Sigma) and cells were harvested by centrifugation at 50g for 5 minutes. Pelleted cells were first resuspended in a PBS wash, harvested by centrifugation, then fixed overnight in ice cold 80% ethanol. Cells were then pelleted by centrifugation at 300g for 5 minutes, washed in PBS and then 1% BSA in PBS. Cells were permeabilized in 0.25% Triton X-100 in PBS for 5 minutes at room temperature, then PBS was added prior to next centrifugation step. Transfected cells (for over-expression experiments) were identified by staining with anti-Flag M2 monoclonal antibody (1:100) for 1 hr at room temperature, followed by 1% BSA in PBS wash and detected by a horse anti-mouse fluorescein antibody (1:30). siRNA transfected cells (down-regulation experiments) were identified by a lack of staining with endogenous anti-hYVH1 antibody. Cells were stained as described above, except non-transfected cells were detected by staining with anti-hYVH1 (1:100) and goat anti-rabbit fluorescein (Vector Labaratories) antibodies. After antibody labelling, cells were washed in 1% BSA in PBS and incubated with RNase A for 30 min prior to addition of propidium iodide (Sigma) for detection of DNA content. As negative and internal controls, unlabelled sample, singly labelled samples and sample labelled only with secondary antibody were included. Two-colour analysis was performed on Cytomics FC500 (Beckman Coulter) using CXP analysis software.

Senescence β-galactosidase assay

Detection of senescent cells was performed according to the protocol provided with a senescence β-galactosidase staining kit (Cell Signaling) with minor modifications.

Briefly, siRNA transfected HeLa cells were washed in PBS and fixed in 2% formaldehyde for 15 minutes at room temperature. After washing, cells were incubated overnight at 37°C in freshly prepared solution of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside in 40mM citric acid/sodium phosphate 6.0, 150mM NaCl, 2mM $MgCl_2$, and 5mM potassium ferricyanide. Cells were visualized on a Zeiss Axiovert 40 CFL microscope in bright field and images were acquired with MicroPublisher 3.3 RTV camera and Q Capture acquisition software.

Results

hYVH1 is regulated by phosphorylation

In order to elucidate whether hYVH1 is regulated by phosphorylation, the total levels of phosphorylation on Flag-YVH1 immunoprecipitated from transiently transfected cells were analyzed by the fluorescent Pro-Q Diamond phosphostain. This fluorescent dye detects phosphorylation on serine, threonine and tyrosine residues, and thus after total protein staining with Coomassie brilliant blue, relative changes in phosphorylation levels can be quantitated by densitometry. Upon treatment of Flag-hYVH1 expressing cells with the PP1 and PP2A inhibitor calyculin A, total phosphorylation was increased about five fold. Furthermore, when Flag-hYVH1 was immunoprecipitated from cells previously treated with calyculin A and then incubated with alkaline phosphatase there was a decrease in the total phosphorylation along with diminished levels of higher molecular weight species (Figure 1A). The identification of the higher molecular weight band as Flag-hYVH1 was confirmed by MS analysis (data not shown). In addition to

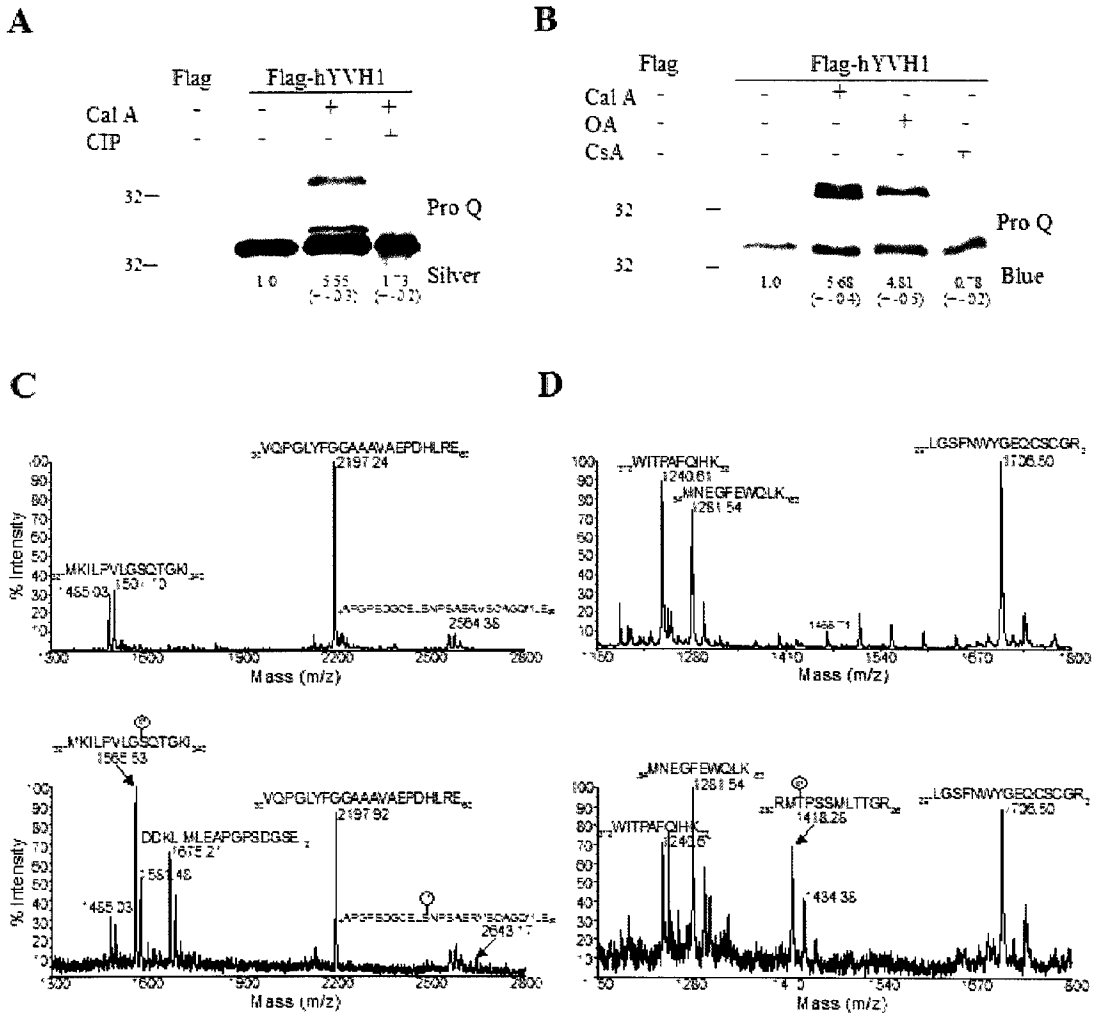


Figure 1. Phosphorylation of hYVH1

(**A**) Flag-tagged hYVH1 was immunoprecipitated from HEK 293 cells and total phosphorylation status was analyzed by staining with Pro-Q Diamond phosphoprotein stain followed by silver staining. 30 minutes prior to lysis, cells were treated with either phosphatase inhibitor calyculin A (Cal A; 100nM) or DMSO (negative control), and afterwards, immunoprecipitated proteins were dephosphorylated by treatment with calf intestinal phosphatase (CIP). (**B**) Transfected cells were treated for 30 minutes prior to lysis with phosphatase inhibitors Cal A, okadaic acid (OA; 5μM) or 24 hr with cyclosporin A (CsA; 10μM). Total phosphorylation status was visualized with Pro-Q diamond stain, followed by Coomassie brilliant blue staining. (**C**, **D**) MS-spectra corresponding to Flag-hYVH1 peptide fingerprint generated with endoproteinase GluC (**C**) and trypsin (**D**). Peptides were obtained from in-gel digestion of immunoprecipitated Flag-hYVH1 treated with Cal A prior to lysis. The upper panel is a representative of a sample prior to phosphopeptide enrichment; the lower panel displays a spectra following the phosphopeptide enrichment step. Arrows indicate phosphopeptides. Peptide identities were confirmed by post-source decay analysis (MS-MS).

calyculin A, treatment with another PP2A phosphatases inhibitor, okadaic acid, caused a similar effect thus confirming our previous result. Consistent with these observations, cyclosporin A, a PP2B inhibitor, did not affect the total phosphorylation status of Flag-hYVH1 (Figure 1B).

In order to identify the sites of phosphorylation, a MS-based approach was employed. Peptides were generated by in-gel digestion with endoproteinase Glu C (Figure 1C) and trypsin (Figure 1D) and analyzed by matrix assisted laser desorption/ionization time of flight (MALDI-TOF) MS. As ionization of phosphopeptides are commonly suppressed by non-phosphopeptides (upper panels), immobilized metal affinity chromatography (IMAC) [23] was carried out to enrich for phosphorylated hYVH1 peptides. IMAC purification followed by MALDI-post source decay analysis allowed us to identify three hYVH1 phosphorylation sites; one within the extreme N-terminal region, corresponding to residue Ser14, one in the extreme C-terminal region, at amino acid position Ser335 and one within the Zn-binding domain, corresponding to Thr252 (Figure 1C-D, lower panels). These results are the first reported phosphorylation sites for any YVH1 orthologue and provide an opportunity to learn a fundamental aspect of hYVH1 function and regulation.

Ser14 phosphorylation affects hYVH1 localization and cell cycle progression

Human YVH1 has been shown to display a nuclear, perinuclear and cytoplasmic localization pattern in a variety of cell lines [9, 16, 18] indicating hYVH1 exhibits nuclear/cytoplasmic shuttling. In order to elucidate a potential role of Ser14 phosphorylation on hYVH1 localization, we examined the localization patterns of non-phosphorylatable S14A and phosphomimetic S14E mutants in HeLa cells (Figure 2). Consistent with previous reports, wild type FLAG-hYVH1 displayed a nuclear,

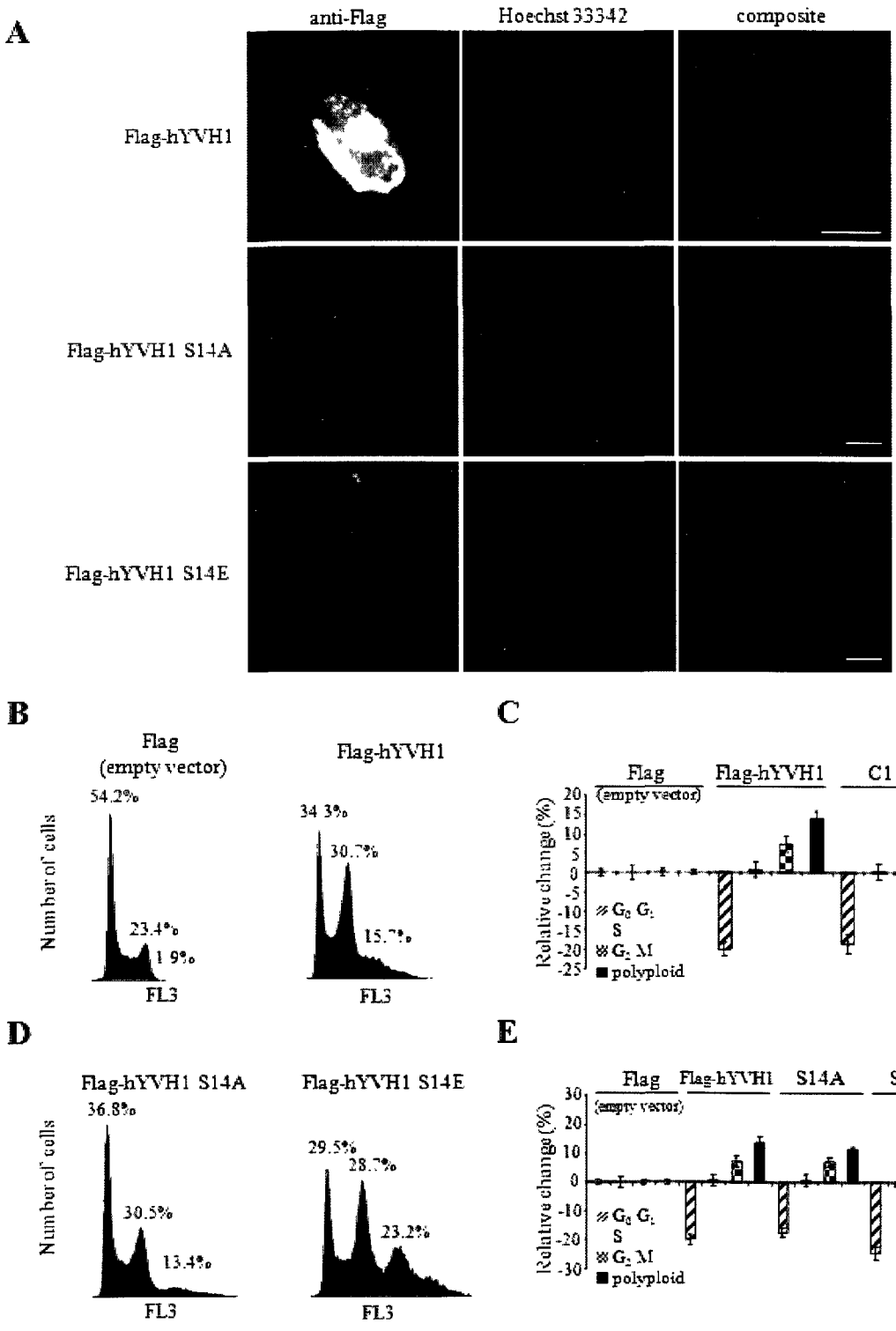

A
anti-Flag
Hoechst 33342
composite
Flag-hYVH1
Flag-hYVH1 S14A
Flag-hYVH1 S14E
B
Flag
(empty vector)
Flag-hYVH1
54.2%
23.4%
1.9%
34.3%
30.7%
15.7%
Number of cells
FL3
C
Flag
Flag-hYVH1
C115S
(empty vector)
Relative change (%)
G0 G1
S
G2 M
polyploid
D
Flag-hYVH1 S14A
Flag-hYVH1 S14E
36.8%
30.5%
13.4%
29.5%
28.7%
23.2%
E
Flag
Flag-hYVH1
S14A
S14E

Figure 2. Analysis of hYVH1 N-terminal phosphorylation site
(**A**) Subcellular localization was analyzed in HeLa cells for Flag-tagged hYVH1 and non-phosphorylatable and phosphomimetic mutants, S14A and S14E, respectively. Transfected cells were detected by anti-Flag antibody staining, and were counterstained with Hoechst 33342. Scale bars represent 15 microns. (**B**) Cell cycle profile of transfected cells presented as representative histogram data plots. HEK 293 cells were transfected with expression vectors encoding either empty vector or Flag-hYVH1 (as indicated), harvested after 48hr, stained for transfected cells, counterstained with propidium iodide and then subjected to a double colour flow cytometry analysis. (**C**) HEK 293 cells were transfected with expression vectors encoding either empty vector, Flag-hYVH1 or Flag-hYVH1 C115S (as indicated), harvested after 48hr, stained for transfected cells, counterstained with propidium iodide and then subjected to a double colour flow cytometry analysis. The graph shows the percentage change in cell cycle phase relative to empty vector control; data shown is a representative of at least three independent experiments. (**D**) Cell cycle profile of cells expressing Ser14 non-phosphorylatable and mimetic mutants presented as representative histogram data plots. Cells were processed as described in B. (**E**) The graph displays a percentage change in cell cycle phase of N-terminal mutants relative to empty vector control; data shown is a representative of at least three independent experiments.

perinuclear and cytoplasmic localization pattern (Figure 2A). It is noted that hYVH1 was not detectable within the nucleolus and that in most instances the shape of the nuclei in transfected cells had a distinct morphology in comparison to neighbouring non-transfected cells (Figure 2A). The S14A mutant displayed a mainly nuclear localization pattern similar to that found with wild-type Flag-hYVH1 but more predominant. As with wild-type over-expression, there were regions lacking staining within the nucleus that were presumed to be nucleoli. The S14E mutation, on the other hand, was predominately localized in the cytoplasm, although some nuclear staining was also detected (Figure 2A).

Unexpectedly, we also noted an overall increase in the number of mitotic-like cells in all hYVH1 variants in comparison to the controls transfected with the Flag-empty vector. In order to further characterize this phenotype, we examined hYVH1 over-expression on cell cycle profiles by flow cytometry utilizing a double colour analysis, in which we gated on cells positive for the presence of individual Flag-hYVH1 constructs. Interestingly, there was a substantial decrease in the percentage of cells in G_0/G_1, an unchanged percentage of cells in S, and a large increase in the proportion in G_2/M and polyploid cells as compared to controls (cells either untransfected or transfected with Flag-empty vector) (Figure 2B). The enzymatically inactive mutant, Flag-hYVH1 C115S, displayed a similar cell cycle profile to Flag-hYVH1 (Figure 2C), thus suggesting that phosphatase activity is not required for hYVH1's effect on the progression through the cell cycle. When the DNA content of cells transfected with the S14A mutant was analyzed, a similar pattern to that for Flag-hYVH1 was observed. However, the mutant mimicking phosphorylation at Ser14 (S14E) caused a more substantial decrease in the proportion of cells in G_0/G_1, and an increase in polyploid cells when evaluated against Flag-hYVH1 (Figure 2D-E). Together, these observations indicate that hYVH1 over-

expression causes an overall increase in 2N DNA content. Furthermore, an N-terminal mutant mimicking the phosphorylated residue, S14E displays a marked increase in cells containing more that 2N DNA content.

Ser335 phosphorylation affects cellular DNA content

In order to assess the functionality of the C-terminal phosphorylation site, non-phosphorylated S335A and phosphomimetic S335E mutations were introduced into Flag-hYVH1 and their effects on hYVH1 localization and cell cycle progression were analyzed. Similar to the S14A mutant, the S335A mutant mainly displayed a nuclear localization pattern with some cytoplasmic staining, while the S335E mutant (like S14E) was predominately present in the cytoplasm (Figure 3A). In addition, over-expression of the phosphomimetic S335E mutant resulted in cells displaying multiple nuclei. This was further confirmed using flow cytometry. Cells overexpressing FlagYVH1-S335E displayed an increase in 2N DNA content *i.e.* an increase in percentage of cells in G_2/M phase, when compared to Flag-hYVH1 expressing cells (Figure 3B-C).

The other phosphorylation site identified within the Zn-binding domain was Thr252. As a proline is immediately following this phosphorylation site, we asked whether proline directed phosphorylation at T252 could be regulated by MAPKs. As a preliminary approach, we tested if various MAPK inhibitors were capable of reducing the levels of phosphorylated hYVH1 *in vivo*. Flag-hYVH1 expressing cells were treated with MAPK inhibitors, SB 203580 (inhibits p38), SP600125 (an inhibitor of JNK) and U0126 (inhibits ERK through inhibition of its upstream activator MEK) alone or in combination with calyculin A. Changes in the phosphorylation status of immunoprecipitated proteins were analyzed by Pro-Q diamond dye. Based on three independent experiments, the total

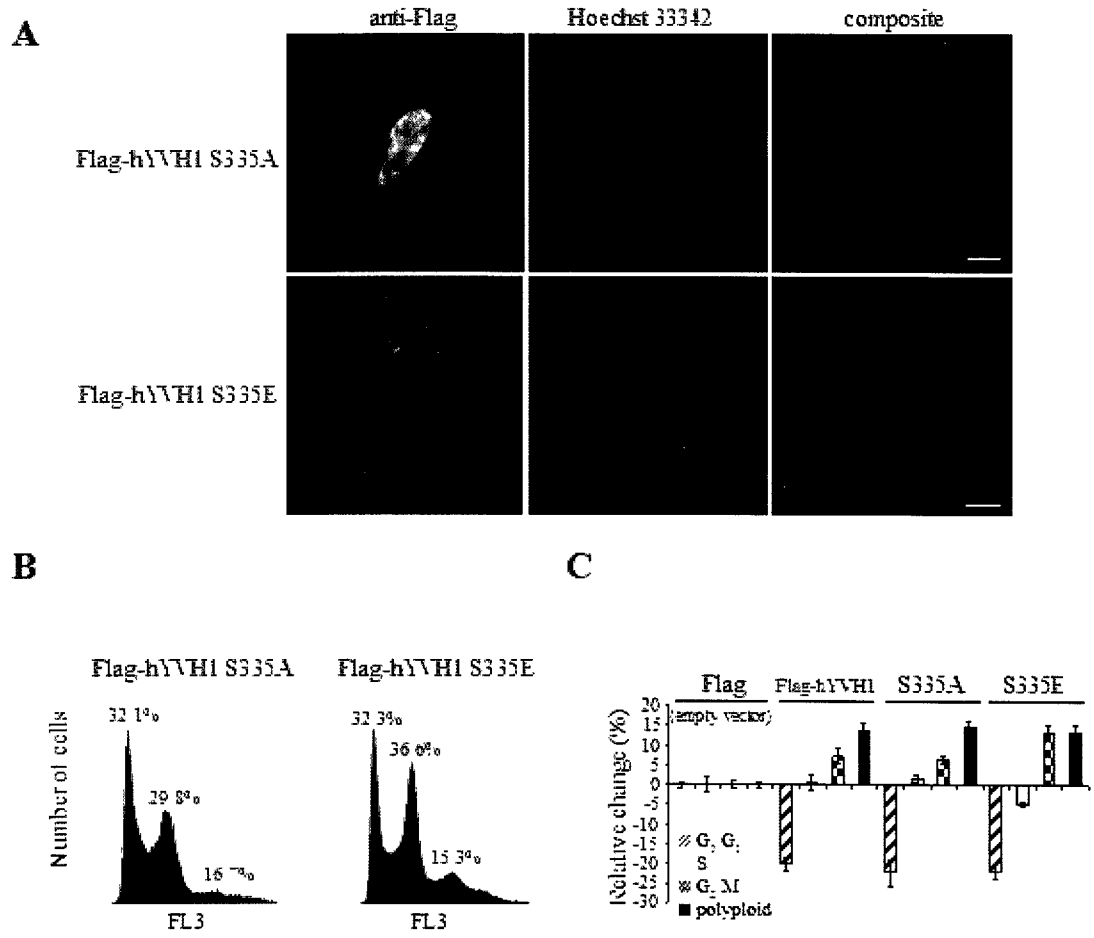


Figure 3. Analysis of hYVH1 C-terminal phosphorylation site

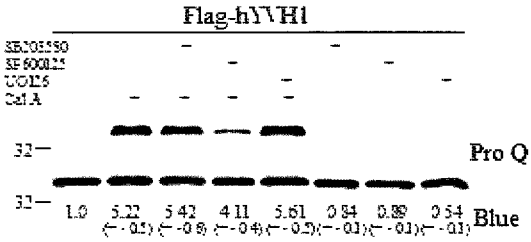
(**A**) Subcellular localization was analyzed in HeLa cells for Flag-tagged hYVH1 and non-phosphorylable and phosphomimetic mutants, S335A and S335E, respectively. Transfected cells were detected by anti-Flag antibody staining, and were counterstained with Hoechst 33342. Scale bars represent 15 microns. (**B**) Cell cycle profile of transfected cells presented as representative histogram data plots. HEK 293 cells were transfected with expression vectors encoding either Flag-hYVH1 S335A or Flag-hYVH1 S335E (as indicated), harvested after 48hr, stained for transfected cells, counterstained with propidium iodide and then subjected to a double colour flow cytometry analysis. (**C**) The graph shows the percentage change in cell cycle phase in cells transfected with C-terminal mutants relative to empty vector control; data shown is a representative of at least three independent experiments.

phosphorylation status was decreased upon combination treatment with SP600125 and calyculin A (Figure 4A). This result suggests that JNK could be one of the upstream regulators of Flag-hYVH1 but requires further investigation. However, this phosphorylation event has no detectable affect on hYVH1 function in terms of localization and ploidy of cells. When the non-phosphorylated T252A and phospho-mimetic T252E mutations were introduced and their ability to affect distribution in the cell by localization were analyzed, in both cases mainly a nuclear pattern was observed (Figure 4B). Regions within the nucleus that lacked staining were tentatively identified as nucleoli. Additionally, the cell cycle profiles of cells expressing these mutants were indistinguishable from cells transfected with Flag-hYVH1 (Figure 4C).

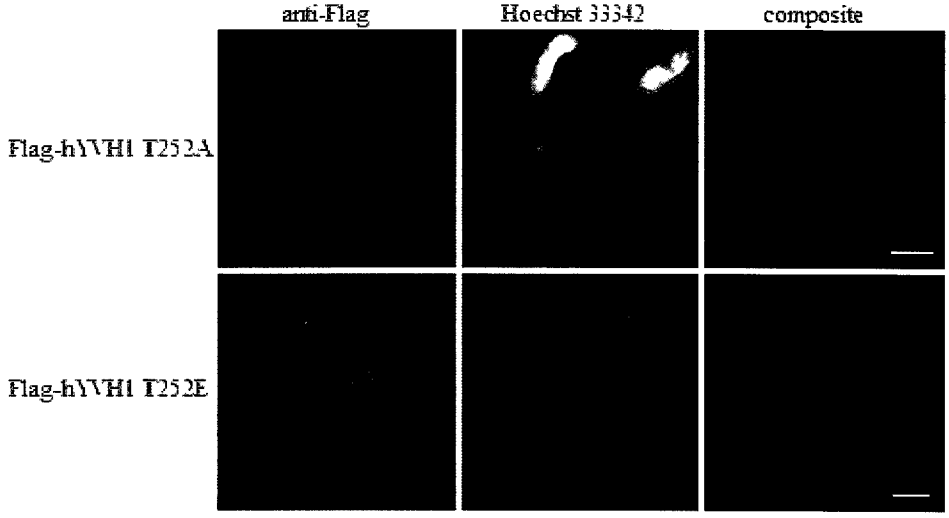
The phosphomimetic mutants, S14E and S335E display localization phenotypes opposite to those observed for the non-phosphorylatable mutants, S14A and S335A. We were next interested in determining whether phosphorylation of these sites acts in a concerted fashion with regards to localization and cell cycle profiles. We therefore examined the effect of double mutant constructs containing both sites mutated at the same time (double AA and double EE mutants) as well as mutants of all three identified sites (triple AAA and triple EEE mutants) on localization (Figure 5A). Not surprisingly, we found a similar distribution to that previously found for the corresponding single mutants with the S14A and S335A mutations mainly nuclear with some cytoplasmic staining and S14E and S335E mutations mainly cytoplasmic.

In contrast, the cell cycle analysis of cells transfected with Flag-hYVH1 S14A S335A (AA) as well as Flag-hYVH1 S14A T252A S335A (AAA) showed less of an effect on the proportion of cells in G_0/G_1 compared to Flag-hYVH1, and on the percentage of cells in G_2/M and multinucleated. The cell cycle profiles were more

A


Flag-hYVH1
SB203580
SP600125
UO126
Cal A
32
Pro Q
32
1.0
5.22
5.42
4.11
5.61
0.84
0.89
0.54
Blue

B


anti-Flag
Hoechst 33342
composite
Flag-hYVH1 T252A
Flag-hYVH1 T252E

C

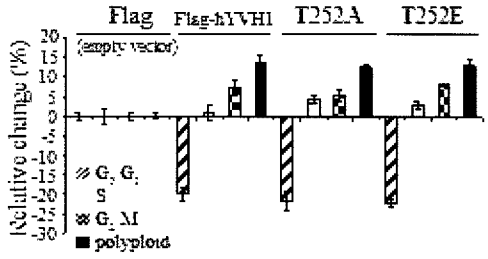

Flag
Flag-hYVH1
T252A
T252E
(empty vector)
Relative change (%)
G1, G2, S
G2, M
polyploid

Figure 4. Analysis of hYVH1 TP phosphorylation site

(**A**) HEK 293 cells were transfected with Flag-hYVH1 and treated for 1hr prior to lysis with MAPK inhibitors SB 203580 (10μM), SP600125 (40μM), U0126 (10μM) without or with calyculin A (Cal A). Total phosphorylation status was visualized with Pro-Q diamond stain, followed by coomassie brilliant blue staining. Densitometry values are presented as ratios of ProQ/coomassie band intensities; DMSO treated Flag-hYVH1 was assigned an arbitrary value of 1.0. Densitometry values were acquired with ImageJ software and are representative of at least three independent experiments. (**B**) Subcellular localization was analyzed in HeLa cells for Flag-tagged hYVH1 and non-phosphorylable and phosphomimetic mutants, T252A and T252E, respectively. Transfected cells were detected by anti-Flag antibody staining, and were counterstained with Hoechst 33342. Scale bars represent 15 microns. (**C**) HEK 293 cells were transfected with expression vectors encoding either Flag-hYVH1 T252A or Flag-hYVH1 T252E (as indicated), harvested after 48hr, stained for transfected cells, counterstained with propidium iodide and then subjected to a double colour flow cytometry analysis. The graph shows the percentage change in cell cycle phase of transfected cells relative to empty vector control; data shown is a representative of at least three independent experiments.

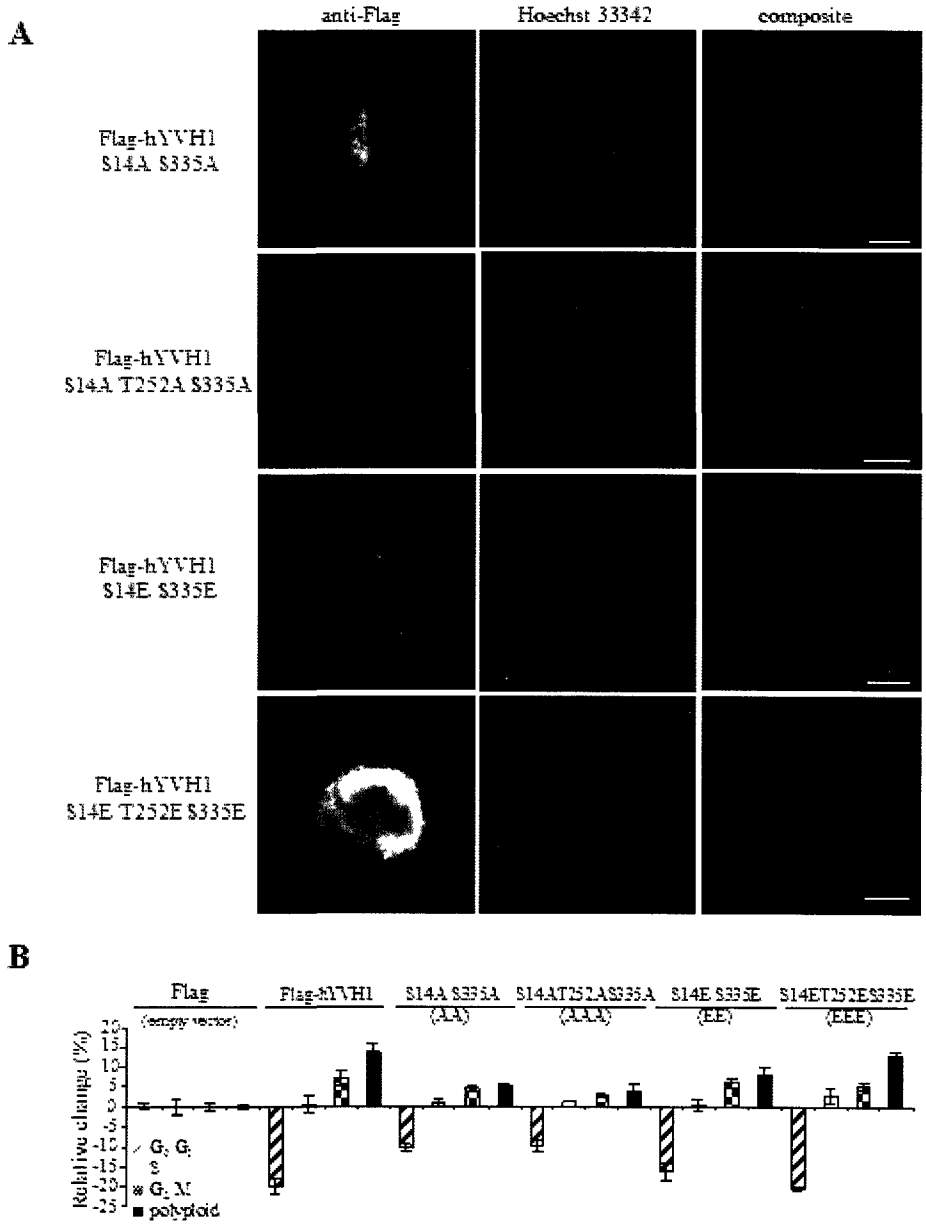

A
anti-Flag
Hoechst 33342
composite
Flag-hYVH1
S14A S335A
Flag-hYVH1
S14A T252A S335A
Flag-hYVH1
S14E S335E
Flag-hYVH1
S14E T252E S335E
B
Flag
(empty vector)
Flag-hYVH1
S14A S335A
(AA)
S14AT252AS335A
(AAA)
S14E S335E
(EE)
S14ET252ES335E
(EEE)
Relative change (%)
G1 G0
S
G2 M
polyploid

Figure 5. Analysis of hYVH1 double and triple phosphorylation site mutants
(**A**) Subcellular localization was analyzed in HeLa cells for Flag-tagged hYVH1 and double and triple non-phosphorylable and phosphomimetic mutants, Flag-hYVH1 S14A S335A (AA), Flag-hYVH1 S14A T252A S335A (AAA) and Flag-hYVH1 S14E S335E (EE) and Flag-hYVH1 S14E T252E S335E (EEE), respectively. Transfected cells were detected by anti-Flag antibody staining, and were counterstained with Hoechst 33342. Scale bars represent 15 microns. (**B**) HEK 293 cells were transfected with expression vectors encoding either Flag-hYVH1 S14A S335A (AA), Flag-hYVH1 S14A T252A S335A (AAA) and Flag-hYVH1 S14E S335E (EE) or Flag-hYVH1 S14E T252E S335E (EEE) (as indicated), harvested after 48hr, stained for transfected cells, counterstained with propidium iodide and then subjected to a double colour flow cytometry analysis. The graph shows the percentage change in cell cycle phase of transfected cells relative to empty vector control; data shown is a representative of at least three independent experiments.

comparable to cell cycle profiles of Flag-empty vector than cells expressing Flag-hYVH1. Thus the triple A mutation abrogates the phenotype caused by over-expression of Flag-hYVH1 (Figure 5B). Meanwhile, the Flag-hYVH1 S14E T252E S335E mutant (EEE) displays a similar effect on the progression through cell cycle as Flag-hYVH1 (Figure 5A).

The zinc binding domain mediates the hYVH1 induced cell cycle profile

In yeast, the C-terminal Zn-binding domain is able to complement the slow-growth phenotype of a *yvh1* deletion [12, 24], rescue the 60S export defect [16, 17, 24], and is necessary for glycogen accumulation, growth and spore maturation [12]. In humans, this domain has a protective function offering intrinsic redox buffering capacity and is necessary for the cytoprotective function of hYVH1 [18, 25]. We therefore tested the over-expression of the Zn-binding domain as well as three different C-terminal deletion constructs for their effect on cell cycle progression. The ΔCT1 construct terminates right after the phosphatase catalytic domain, the ΔCT2 variant in addition to the catalytic domain contains the first zinc coordination site, and the ΔCT3 variant contains an additional hydrophobic motif, which is highly conserved among orthologues [18].

The N-terminal catalytic domain (ΔCT1) essentially had a similar profile as our Flag-empty vector control, indicating the importance of the zinc binding domain for the hYVH1 cell cycle phenotype. Interestingly, over-expression of both partial (ΔCT2 and ΔCT3) and the complete C-terminal Zn-binding domain resulted in a large decrease in the population of G_0/G_1 cells and a large increase in the population of G_2/M and polyploid cells (Figure 6A-B). The strongest effect was found with the entire Zn-binding domain.

These results strongly suggest that the Zn-binding domain is necessary and sufficient for hYVH1 mediated cell cycle changes.

Knockdown of hYVH1 expression blocks cells in G_0/G_1 and induces cellular senescence

The above experiments demonstrate that over-expression of hYVH1 reduces G_0/G_1 and increases the percentage of cells in G_2/M and the percentage of cells displaying DNA content greater than 2N. To complement these results, we analyzed the cell cycle profile of cells that had hYVH1 expression knocked down by double stranded small interfering RNA (siRNA). The hYVH1 siRNA oligonucleotide was transfected into both HeLa and HEK293 cells and resulted in approximately an 80% decrease in hYVH1 expression, while actin levels remained unchanged (Figure 7E). The loss of hYVH1 expression was accompanied by a striking increase in the G_0/G_1 population of cells and subsequent decrease in other cell cycle phases (Figure 7A-B). When we gated on cells that were not expressing hYVH1, approximately 90% of the cells were in G_0/G_1 (Figure 7A).

Since cells that arrest in G_0/G_1 for prolonged periods are susceptible to permanent cell cycle arrest, we were interested in investigating if hYVH1 can influence cellular senescence. Therefore, cells treated with hYVH1 siRNA were stained for the senescent marker β-galactosidase and analyzed at different time periods (Figure 7C). Staining of hYVH1 siRNA treated cells revealed that post seven days siRNA there was a reproducible increase in the proportion of senescent cells (Figure 7C-D). Coupled to the over-expression data, these results demonstrate for the first time, that hYVH1 is a novel regulator of cell cycle progression.

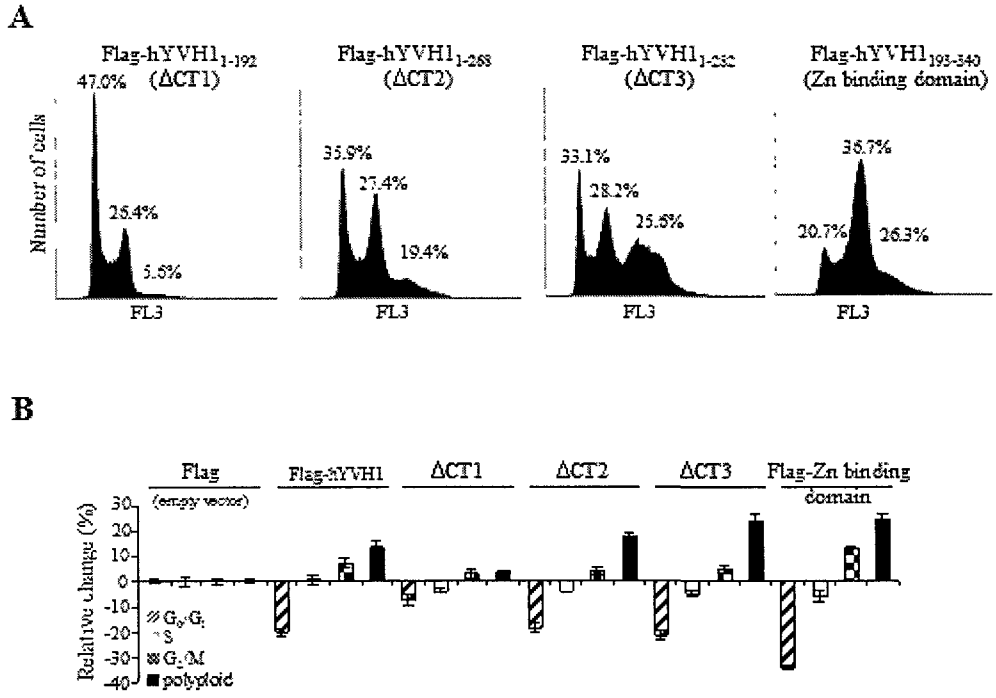


Figure 6. Effect of domain specific constructs on progression through cell cycle
(**A**) Cell cycle profile of transfected cells presented as representative histogram data plots. HEK 293 cells were transfected with expression vectors encoding either Flag-hYVH1$_{1-192}$ (ΔCT1), Flag-hYVH1$_{1-268}$ (ΔCT2), Flag-hYVH1$_{1-282}$ (ΔCT3), or Flag-hYVH1$_{193-340}$ (Zn-binding domain) (as indicated), harvested after 48hr, stained for transfected cells, counterstained with propidium iodide and then subjected to a double colour flow analysis. (**B**) The graph shows the percentage change in cell cycle phase in cells transfected with various domains relative to empty vector control; data shown is a representative of at least three independent experiments.

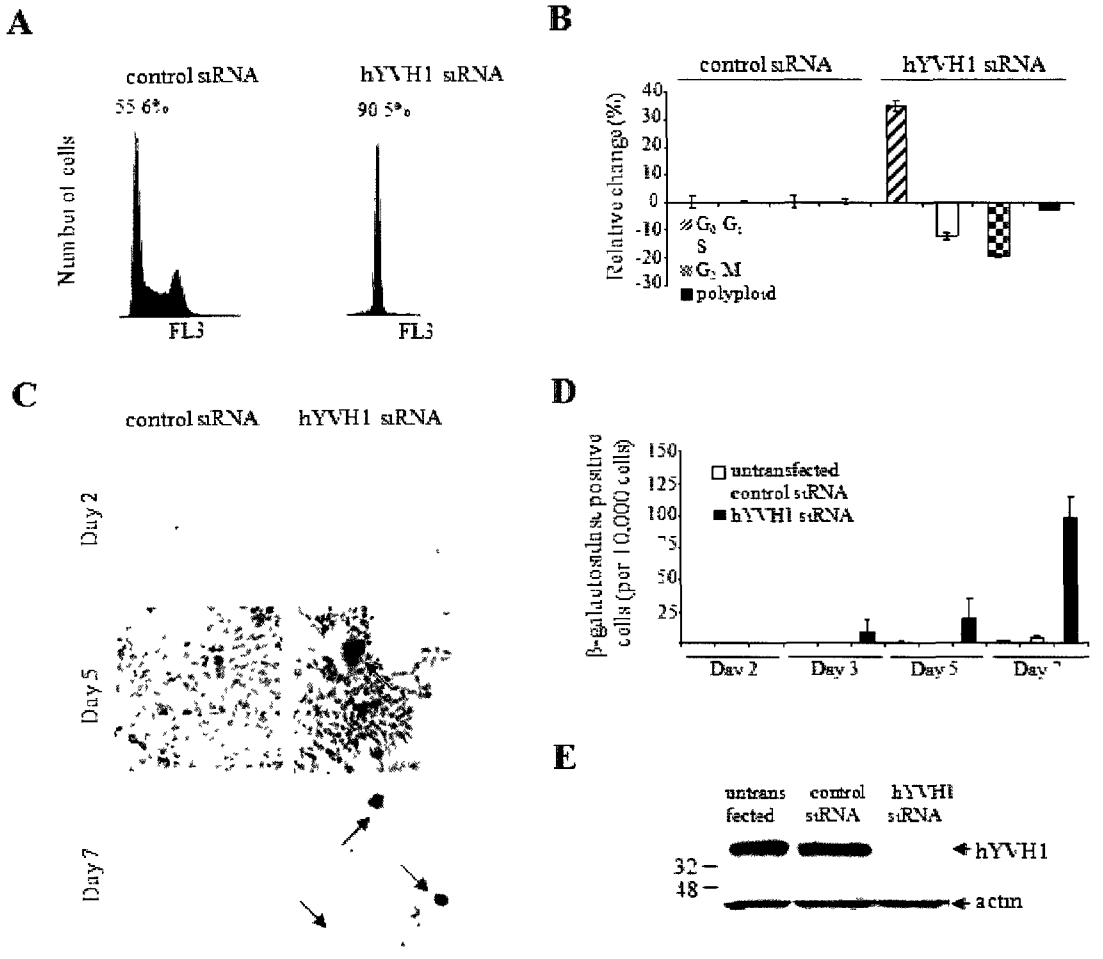


Figure 7. Analysis of hYVH1 siRNA

(**A**) Cell cycle profile of control siRNA treated cells or cells where hYVH1 has been down-regulated by siRNA presented as representative histogram data plots. HEK 293 cells were transfected with either control siRNA or hYVH1 siRNA (as indicated), harvested after 48hr, stained for non-transfected cells (lack of staining with endogenous hYVH1 antibody), counterstained with propidium iodide and then subjected to a double colour flow cytometry analysis. (**B**) The graph shows the percentage change in cell cycle phase in cells where hYVH1 has been down-regulated by siRNA relative to control siRNA; data shown is a representative of at least three independent experiments. (**C**) β-galactosidase staining of HEK 293 cells transfected with a control siRNA or hYVH1 siRNA. Cells were cultured for two, five and seven days. Arrows indicate positive blue cells. (**D**) Graphical presentation of the β-galactosidase staining of senescent cells. Data shown is a representative of three independent experiments. (**E**) Anti-hYVH1 immunoblot analysis of untransfected cells, cells transfected with siRNA control and hYVH1 siRNA and corresponding anti-actin immunoblot.

Discussion

We have discovered that the expression of hYVH1 alters the cell cycle profile of asynchronized HEK 293 cells. While the phosphatase activity seems to be dispensable for this property, the zinc binding domain is necessary and sufficient to elicit this response. Although this study is the first report to implicate hYVH1 as a cell cycle regulator, other atypical DUSPs have been suggested to act as important positive and negative regulators for cell cycle progression. For example, expression of VHR (DUSP3), has been shown to be most abundant in mitotic cells [26]. Down-regulation of VHR is accompanied by loss of entry into S phase, thus triggering cell cycle arrest [26]. Down-regulation of laforin results in a significant increase in the number of cells at S and G_2/M phases, thus promoting cell growth [27], Similarly, down-regulation of DUSP11, also results in increased proliferation [28].

IMAC purification coupled to MALDI-TOF MS yielded the first ever identified phosphorylation sites for any YVH1 orthologue. The two phosphorylation sites that affected hYVH1 localization and cell cycle progression are located at the termini of hYVH1. The location of the phosphorylation sites is not surprising since the termini of proteins is a common site for protein phosphorylation including cell cycle regulators p53 [29], histone H3 [30] and cdc2 [31]. Since phosphorylation often regulates enzyme activity, we did initially examine if these sites affected phosphatase activity. However, we saw no difference in phosphatase activity upon treatment with calyculin A using the exogenous fluorescent substrate DiFMUP (data not shown). Instead, our results suggest that both the N-terminal Ser14 and C-terminal Ser335 phosphorylation sites play a role in

regulating the subcellular localization of hYVH1 and in augmenting the cell cycle phenotype exhibited by hYVH1 over-expression.

Analysis of the conservation of the phosphorylation sites amongst YVH1 orthologues suggests that regulatory phosphorylation may be specific for mammalian orthologues. In particular, the Ser14 phosphorylation site within the N-terminal region of hYVH1 is not conserved in yeast or rodents but is present in some advanced vertebrates including the primate orthologues, suggesting that this site has arisen more recently and may induce novel regulatory control on hYVH1 function. The Ser335 phosphorylation site on the other hand is extensively conserved in mammals (including rodents) but is absent in yeast. Furthermore, the Ser335 phosphorylation site located five amino acids from the C-terminus resides in a so called SQ site. This consensus sequence is targeted by members of the phosphatidyl-inositol-3-kinase protein kinase family, including ATM (ataxia telangiectasia mutated), ATR (ATM and Rad3 related) and DNA-PK (DNA-protein kinase) [32]. Interestingly, under normal cell culture conditions, ATM may become activated by calyculin A or okadaic acid treatment without the need for DNA damage [33]. This activity is cell cycle dependent and at a maximum level during G_2/M [34] and upon nocodazole and taxol induced mitotic stress [35]. Thus since both okadaic acid and calyculinA treatment increased hYVH1 phosphorylation, and hYVH1 over-expression increased the percentage of cells in G_2/M, it will be important to examine if Ser335 phosphorylation occurs in a G_2/M specific manner. Moreover, we are interested in testing if ATM is the upstream kinase responsible for hYVH1 phosphorylation at Ser335.

Our preliminary experiments demonstrated that a JNK inhibitor reduced the phosphorylation of hYVH1. The third phoshorylation site identified, Thr252, located within the Zn-binding domain, is a proline directed site, the minimum consensus site for

JNK. Also, hYVH1 has been shown to play a role in protecting cells from various cellular stresses including hydrogen peroxide, heat stress, and Fas receptor activation [18]. Since these cellular insults are known JNK activators [36] we are particularly interested in examining if JNK can directly phosphorylate Thr252.

A large scale experiment targeted at identification of subunits involved in yeast ribosomal biogenesis (60S pre-ribosomal particles), identified YVH1 from Arx1- and Kre35-TAP tag pull down screens by tandem mass spectrometry [15]. As Arx1 displays nucleoplasmic and also cytoplasmic localization and Kre35 exhibits cytoplasmic distribution with nuclear exclusion [15], this is in accordance with herein reported distribution of Flag-hYVH1. In the 60S ribosomal biogenesis pathway, the early pre-60S particles are restricted to the nucleolus, the later particles are released into the nucleoplasm and then exported into cytoplasm [37]. As the nonphosphorylatable N- and C-terminal mutants displayed predominantly nuclear localization pattern, this suggests a possibility that the phosphorylation event could be the underlying mechanism required for loading onto pre-60S prior to nuclear export, thus controlling one of the late maturation steps in the 60S biogenesis pathway. We are currently investigating if the hYVH1 phosphorylation mutants affect 60S subunit association and the rate of ribosome biogenesis.

One of the characteristic features observed upon over-expression of the phosphorylation site mutants was an aberrant nuclear morphology phenotype accompanied by an increased proportion of cells with an increased DNA content as detected by flow cytometry. Moreover, knocking down hYVH1 expression resulted in the opposite cell cycle phenotype (increase in G_0/G_1) along with an increase in cellular senescence. Deletion of YVH1 in yeast has been associated with slow-growth phenotype

[9, 11, 12, 16, 17, 24] and a delay in nuclear pre-ribosomal RNA processing, through impaired nuclear export of 60S subunit [16, 17, 24]. These observations suggest that hYVH1 over-expression may contribute to genomic instability, a hallmark of cancer. Consistent with this, an increase in DNA copy numbers of the physical location of human hYVH1 at 1q [9], has been observed in numerous cancers, including retinoblastoma [38], liposarcomas [39], and intracranial ependymoma [40]. hYVH1 was proposed to be oncogenic in the latter subset of cancers [40].

The exact mechanism by which hYVH1 affects DNA content is still unclear. One possibility is that hYVH1 over-expression forces endoduplication, where cells undergo repeated rounds of genomic DNA synthesis without an intervening mitosis, producing polyploid cells [41]. Alternatively, the increased DNA content may reflect a failure of cytokinesis or uncontrolled DNA replication. It is tempting to speculate, that highly regulated hYVH1 expression is necessary for normal progression through cell cycle, as both, over-expression and down-regulation, affect cell cycle profiles.

Acknowledgements

We thank members of P.O.V and J.W.H laboratories for discussions and critical comments. This work was funded by Discovery Grants from Natural Sciences and Engineering Research Council (NSERC) of Canada to P.O.V and J.W.H and NSERC post-graduate scholarship to A.K.

References

1. Hunter, T., *Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling.* Cell, 1995. **80**(2): p. 225-36.

2. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

3. Virshup, D.M. and S. Shenolikar, *From promiscuity to precision: protein phosphatases get a makeover.* Mol Cell, 2009. **33**(5): p. 537-45.

4. Farooq, A. and M.M. Zhou, *Structure and regulation of MAPK phosphatases.* Cell Signal, 2004. **16**(7): p. 769-79.

5. Patterson, K.I., et al., *Dual-specificity phosphatases: critical regulators with diverse cellular targets.* Biochem J, 2009. **418**(3): p. 475-89.

6. Wang, J., et al., *A unique carbohydrate binding domain targets the lafora disease phosphatase to glycogen.* J Biol Chem, 2002. **277**(4): p. 2377-80.

7. Yuan, Y., D.M. Li, and H. Sun, *PIR1, a novel phosphatase that exhibits high affinity to RNA . ribonucleoprotein complexes.* J Biol Chem, 1998. **273**(32): p. 20347-53.

8. Beeser, A.E. and T.G. Cooper, *The dual-specificity protein phosphatase Yvh1p acts upstream of the protein kinase mck1p in promoting spore development in Saccharomyces cerevisiae.* J Bacteriol, 1999. **181**(17): p. 5219-24.

9. Muda, M., et al., *Identification of the human YVH1 protein-tyrosine phosphatase orthologue reveals a novel zinc binding domain essential for in vivo function.* J Biol Chem, 1999. **274**(34): p. 23991-5.

10. Roma-Mateo, C., et al., *A novel phosphatase family, structurally related to dual-specificity phosphatases, that displays unique amino acid sequence and substrate specificity.* J Mol Biol, 2007. **374**(4): p. 899-909.

11. Guan, K., et al., *A yeast protein phosphatase related to the vaccinia virus VH1 phosphatase is induced by nitrogen starvation.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 12175-9.

12. Beeser, A.E. and T.G. Cooper, *The dual-specificity protein phosphatase Yvh1p regulates sporulation, growth, and glycogen accumulation independently of catalytic activity in Saccharomyces cerevisiae via the cyclic AMP-dependent protein kinase cascade.* J Bacteriol, 2000. **182**(12): p. 3517-28.

13. Sakumoto, N., et al., *Dual-specificity protein phosphatase Yvh1p, which is required for vegetative growth and sporulation, interacts with yeast pescadillo*

homolog in Saccharomyces cerevisiae. Biochem Biophys Res Commun, 2001. **289**(2): p. 608-15.

14. Oeffinger, M., et al., *Yeast Pescadillo is required for multiple activities during 60S ribosomal subunit synthesis.* Rna, 2002. **8**(5): p. 626-36.

15. Nissan, T.A., et al., *60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm.* Embo J, 2002. **21**(20): p. 5539-47.

16. Lo, K.Y., et al., *Ribosome stalk assembly requires the dual-specificity phosphatase Yvh1 for the exchange of Mrt4 with P0.* J Cell Biol, 2009. **186**(6): p. 849-62.

17. Kemmler, S., et al., *Yvh1 is required for a late maturation step in the 60S biogenesis pathway.* J Cell Biol, 2009. **186**(6): p. 863-80.

18. Sharda, P.R., et al., *The dual-specificity phosphatase hYVH1 interacts with Hsp70 and prevents heat-shock-induced cell death.* Biochem J, 2009. **418**(2): p. 391-401.

19. Biernacki, M.A., et al., *Efficacious immune therapy in chronic myelogenous leukemia (CML) recognizes antigens that are expressed on CML progenitor cells.* Cancer Res, 2010. **70**(3): p. 906-15.

20. Jordan, M., A. Schallhorn, and F.M. Wurm, *Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation.* Nucleic Acids Res, 1996. **24**(4): p. 596-601.

21. Blum, H., H. Beier, and H.J. Gross, *Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels.* Electrophoresis, 1987. **8**(2): p. 93-99.

22. Sharon, M.A., et al., *Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp Artemia franciscana.* Biochem Cell Biol, 2009. **87**(2): p. 415-30.

23. Zhou, W., et al., *Detection and sequencing of phosphopeptides affinity bound to immobilized metal ion beads by matrix-assisted laser desorption/ionization mass spectrometry.* J Am Soc Mass Spectrom, 2000. **11**(4): p. 273-82.

24. Liu, Y. and A. Chang, *A mutant plasma membrane protein is stabilized upon loss of Yvh1, a novel ribosome assembly factor.* Genetics, 2009. **181**(3): p. 907-15.

25. Bonham, C.A. and P.O. Vacratsis, *Redox regulation of the human dual specificity phosphatase YVH1 through disulfide bond formation.* J Biol Chem, 2009. **284**(34): p. 22853-64.

26. Rahmouni, S., et al., *Loss of the VHR dual-specific phosphatase causes cell-cycle arrest and senescence.* Nat Cell Biol, 2006. **8**(5): p. 524-31.

27. Liu, R., et al., *Laforin negatively regulates cell cycle progression through glycogen synthase kinase 3beta-dependent mechanisms.* Mol Cell Biol, 2008. **28**(23): p. 7236-44.

28. Caprara, G., et al., *Isolation and characterization of DUSP11, a novel p53 target gene.* J Cell Mol Med, 2009. **13**(8B): p. 2158-70.

29. Amano, T., et al., *Simultaneous phosphorylation of p53 at serine 15 and 20 induces apoptosis in human glioma cells by increasing expression of pro-apoptotic genes.* J Neurooncol, 2009. **92**(3): p. 357-71.

30. Dyson, M.H., et al., *MAP kinase-mediated phosphorylation of distinct pools of histone H3 at S10 or S28 via mitogen- and stress-activated kinase 1/2.* J Cell Sci, 2005. **118**(Pt 10): p. 2247-59.

31. Berry, L.D. and K.L. Gould, *Regulation of Cdc2 activity by phosphorylation at T14/Y15.* Prog Cell Cycle Res, 1996. **2**: p. 99-105.

32. Traven, A. and J. Heierhorst, *SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins.* Bioessays, 2005. **27**(4): p. 397-407.

33. Huang, X., et al., *Sequential phosphorylation of Ser-10 on histone H3 and ser-139 on histone H2AX and ATM activation during premature chromosome condensation: relationship to cell-cycle phase and apoptosis.* Cytometry A, 2006. **69**(4): p. 222-9.

34. McManus, K.J. and M.J. Hendzel, *ATM-dependent DNA damage-independent mitotic phosphorylation of H2AX in normally growing mammalian cells.* Mol Biol Cell, 2005. **16**(10): p. 5013-25.

35. Shen, K., et al., *ATM is activated by mitotic stress and suppresses centrosome amplification in primary but not in tumor cells.* J Cell Biochem, 2006. **99**(5): p. 1267-74.

36. Bode, A.M. and Z. Dong, *The functional contrariety of JNK.* Mol Carcinog, 2007. **46**(8): p. 591-8.

37. Tschochner, H. and E. Hurt, *Pre-ribosomes on the road from the nucleolus to the cytoplasm.* Trends Cell Biol, 2003. **13**(5): p. 255-63.

38. Gratias, S., et al., *Genomic gains on chromosome 1q in retinoblastoma: consequences on gene expression and association with clinical manifestation.* Int J Cancer, 2005. **116**(4): p. 555-63.

39. Kresse, S.H., et al., *Mapping and characterization of the amplicon near APOA2 in 1q23 in human sarcomas by FISH and array CGH.* Mol Cancer, 2005. **4**: p. 39.

40. Mendrzyk, F., et al., *Identification of gains on 1q and epidermal growth factor receptor overexpression as independent prognostic markers in intracranial ependymoma.* Clin Cancer Res, 2006. **12**(7 Pt 1): p. 2070-9.

41. Sivaprasad, U., Y.J. Machida, and A. Dutta, *APC/C--the master controller of origin licensing?* Cell Div, 2007. **2**: p. 8.

CHAPTER 5

Conclusions and Future Directions

In conclusion, the work presented here displays the applicability of MS-based approaches to various biological systems. This powerful analytical tool was used not only to identify a novel protein from *Artemia franciscana*, a model organism for which the genome sequence is not available (Chapter 2), analyze cysteine modifications on protein disulfide isomerase due to changes in the redox state (Chapter 3), but also to discover hYVH1 residues modified by phosphorylation (Chapter 4). In each of these cases the data generated by mass spectrometry provided a valuable and initial step in the quest for the functional significance of these proteins.

The capability of MS-based approaches to *de novo* identify a 21 kDA protein from *A. franciscana* diapaused embryos as a group 1 LEA protein [1] identified for the first time the presence of group 1 LEA proteins in animals. Until our discovery, the group 1 LEA proteins were only described for non-animal species [2, 3]. Our observations also support a role for the group 1 LEA protein in desiccation tolerance in *A. franciscana* encysted embryos [1]. As encysted embryos of brine shrimp *A. franciscana* are one of the most resistant organisms to withstand extreme environmental conditions, including the almost complete loss of water [4], our work sheds further light onto molecular mechanisms governing this process. Furthermore, the sequence assignment to the 21 kDa protein allowed the generation of antibodies, which have led to the identification of multiple group 1 LEA proteins in *A. franciscana* encysted embryos. We have recently submitted an additional manuscript in this regard.

A combination of an *in vitro* isotopic labelling technique and MS-based analysis was employed in order to assess relative changes in the dynamic thiol and disulfide states of recombinant human protein disulfide isomerase, PDI [5]. ICAT labelling allowed the monitoring of changes in the redox states of individual cysteine residues within PDI

under auto-oxidation conditions. As PDI contains four cysteines within two active sites and two cysteines within the so-called *b'* domain, ICAT labelling of the active site cysteines supported previously suggested non-equivalency between these active sites [6, 7]. As the possibility exists, that different environmental stress conditions affect the redox status and thus activity of PDI, the approach described with minor modifications would be an invaluable tool in future experiments.

As the major focus of Dr. Vacratsis' laboratory is the study of atypical dual specificity phosphatases, particularly hYVH1, a more detailed overview of future experiments regarding this protein is provided.

To our knowledge, the work described herein on the identification and characterization of hYVH1 (DUSP12) phosphorylation sites, is novel for this family of proteins. Identification of these sites provides a basis for several lines of future investigation. For example, one of the next rational steps is to generate phosphospecific antibodies, which will further facilitate characterization of the functional significance of these sites. Additionally, cell cycle synchronization experiments, achieved upon drug treatments, such as double thymidine block for G_1/S, mimosine for G_1 and nocodazole for M phase [8] or by release from double thymidine block, will be a useful approach to address phosphorylation dependent changes in the localization pattern of hYVH1 at the endogenous level. As hYVH1 displays nuclear, perinuclear and cytoplasmic localization [9-11] a possibility exists, that these localization patterns are cell cycle dependent. Furthermore, as the identified phosphorylation sites affect the progression through cell cycle, the use of phosphospecific antibodies will allow the detection of the sub-population of cells that are phosphorylated at these specific sites by the analysis of their DNA content by flow cytometry. Last, but not least, the possibility exists, that the

phosphorylation at these sites might be upregulated in response to stress (and downregulated by treatment with specific inhibitors) or differentially expressed in different cell lines. In order to address these possibilities, phosphospecific antibodies are valuable tools that will allow the determination of relative phosphorylation levels by western blot analysis of total cellular lysates from cells under various conditions. hYVH1 has also been shown to be overexpressed in chronic myelogenous leukemia [12] and amplification of the hYVH1 gene region has been detected in numerous cancers [13-15]. Thus the use of phosphospecific antibodies in analysis may provide clues as to whether phosphorylation of these sites is correlated with disease. Furthermore, our results suggest that several of these sites are important for the generation of DNA content changes in cells upon hYVH1 over-expression. The question is, do they act similarly in malignancies? This may be in part accomplished by an immunohistochemistry screen of commercially available core patient biopsies attached onto a microscope slide.

One of the limitations associated with phosphospecific antibodies is their insufficient performance in immunoprecipitation assays. As there is the possibility, that phenotypes associated with over-expression of the phosphomutants are caused either by novel protein-protein interactions or disruption of a critical protein-protein complexes, a MS-based TAP-tag approach would allow the identification of novel interacting partners. The TAP-tag method is efficient at the identification of interacting partners at sub-stoichiometric levels [16] and was introduced in detail in Chapter 1. Since over-expression of hYVH1 has shown to result in an increased DNA content, wild-type hYVH1 and phosphomimetic S14E and S335E mutants expressed from a TAP-tag vector will be analyzed in unsynchronized and synchronized cells, as the over-expression has shown an increase in DNA content. Alternatively, to address the possibility that any of

the potential interacting partners are hYVH1 substrates, an additional mutation of catalytic aspartic acid to alanine (D84A) will be introduced into these phosphomutants thus modifying them into "substrate-traps". The "substrate-trap" mutation impedes the substrate dephosphorylation process. It is noted that while the phosphatase-substrate complex is still formed, the subsequent cleavage of the thiol phosphate bond initiated by the nucleophilic cysteine cannot take place without the catalytic acid/base (Asp84), thus rendering the substrate attached and stable enough to withstand isolation [17]. In order to reduce the complexity of immunoprecipitated proteins, separation will be performed either by SDS-PAGE or MudPIT-base method [18] (and introduced in detail in Chapter 1). The latter option has recently become a feasible alternative with the acquisition of a nano-spray triple quadrupole tandem MS instrument in our laboratory.

The three dimensional structure of proteins, experimentally elucidated either from x-ray crystallography or NMR studies, provide invaluable insight on how structure dictates functions. To date, crystal structures of only three atypical DUSPs have been elucidated, including VHR (DUSP3) [19], JKAP (DUSP22) [20] and VHZ (DUSP23) [21]. All of these are relatively small proteins and contain only the catalytic domain. As hYVH1 is the only DUSP that contains a phosphatase catalytic domain and a zinc binding domain, this further complicates calculations for the computational homology model. Dr. Vacratsis' laboratory is currently in the process of elucidating the x-ray structure of hYVH1 (DUSP12) in collaboration with Dr. B. Crane (Cornell University). One of the unpredictable properties of proteins is their refractory nature to crystallization, requiring a long process of optimization. Thus in the meantime, MS-based methods such as the cross-linker MS approach can identify peptides that are spatially close in a folded protein.

Cross-linked peptides identified from MS-analysis of isotopically labelled cross-linked protein are indicative of the spatial proximity of two linked amino acids at the time of cross-linking [22, 23]. Upon labelling of intact proteins with a mixture of isotopically labelled 'heavy' disuccinimidyl suberate (DSS) containing 12 deuteriums (D_{12}DSS) [23] and commercially available 'light' D_0DSS, cross-linker-modified peptides are detected by presence of isotopic shift between labelled fragment ion peaks. Experimental evidence indicated, that DSS formed peptide-peptide links connected mostly lysine residues [22]. After cross-linking the amino acids within the protein and prior to proteolysis, disulfide bonds were reduced, and thiol groups were alkylated with iodoacetamide. The use of trypsin to generate peptide fragments has two advantages; first, cleavage does not occur at the lysine residues participating in the cross-links, thus this miscleavage is included as a pre-selection criteria; and second, the ability to enrich cross-linked peptide by strong cation exchange chromatography. Fractionation steps prior to MS-analysis are necessary, especially in a complex sample, as the cross-linked peptides are present at low stoichiometry and occur at low frequency compared to unmodified peptides [22]. This method is applicable to both MALDI [23] and electrospray ionization [22] based MS-instruments. With the latter instrumental set up, the majority of detected cross-linked peptides have a charge state greater than +3 [22], thus this property can be incorporated into criteria for inclusion list of peptides selected for MS/MS analysis.

hYVH1 contains eleven lysine residues within the catalytic domain and nine lysine residues within the zinc binding domain, thus with this MS-based approach the interaction interface between these two domains should be identifiable. If the sensitivity and detection power of our MALDI-TOF instrument is insufficient for the proposed

experiments, a nano-spray triple quadrupole mass spectrometer, the newest addition to our department will be invaluable.

References

1. Sharon, M.A., et al., *Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp Artemia franciscana.* Biochem Cell Biol, 2009. **87**(2): p. 415-30.

2. Tunnacliffe, A. and M.J. Wise, *The continuing conundrum of the LEA proteins.* Naturwissenschaften, 2007. **94**(10): p. 791-812.

3. Battaglia, M., et al., *The enigmatic LEA proteins and other hydrophilins.* Plant Physiol, 2008. **148**(1): p. 6-24.

4. Clegg, J.S., *Protein stability in Artemia embryos during prolonged anoxia.* Biol Bull, 2007. **212**(1): p. 74-81.

5. Kozarova, A., et al., *Identification of redox sensitive thiols of protein disulfide isomerase using isotope coded affinity technology and mass spectrometry.* J Am Soc Mass Spectrom, 2007. **18**(2): p. 260-9.

6. Lyles, M.M. and H.F. Gilbert, *Mutations in the thioredoxin sites of protein disulfide isomerase reveal functional nonequivalence of the N- and C-terminal domains.* J Biol Chem, 1994. **269**(49): p. 30946-52.

7. Darby, N.J., J. Kemmink, and T.E. Creighton, *Identifying and characterizing a structural domain of protein disulfide isomerase.* Biochemistry, 1996. **35**(32): p. 10517-28.

8. Harper, J.V., *Synchronization of cell populations in G1/S and G2/M phases of the cell cycle.* Methods Mol Biol, 2005. **296**: p. 157-66.

9. Lo, K.Y., et al., *Ribosome stalk assembly requires the dual-specificity phosphatase Yvh1 for the exchange of Mrt4 with P0.* J Cell Biol, 2009. **186**(6): p. 849-62.

10. Sharda, P.R., et al., *The dual-specificity phosphatase hYVH1 interacts with Hsp70 and prevents heat-shock-induced cell death.* Biochem J, 2009. **418**(2): p. 391-401.

11. Muda, M., et al., *Identification of the human YVH1 protein-tyrosine phosphatase orthologue reveals a novel zinc binding domain essential for in vivo function.* J Biol Chem, 1999. **274**(34): p. 23991-5.

12. Biernacki, M.A., et al., *Efficacious immune therapy in chronic myelogenous leukemia (CML) recognizes antigens that are expressed on CML progenitor cells.* Cancer Res, 2010. **70**(3): p. 906-15.

13. Mendrzyk, F., et al., *Identification of gains on 1q and epidermal growth factor receptor overexpression as independent prognostic markers in intracranial ependymoma.* Clin Cancer Res, 2006. **12**(7 Pt 1): p. 2070-9.

14. Gratias, S., et al., *Genomic gains on chromosome 1q in retinoblastoma: consequences on gene expression and association with clinical manifestation.* Int J Cancer, 2005. **116**(4): p. 555-63.

15. Kresse, S.H., et al., *Mapping and characterization of the amplicon near APOA2 in 1q23 in human sarcomas by FISH and array CGH.* Mol Cancer, 2005. **4**: p. 39.

16. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.

17. Flint, A.J., et al., *Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases.* Proc Natl Acad Sci U S A, 1997. **94**(5): p. 1680-5.

18. Link, A.J., et al., *Direct analysis of protein complexes using mass spectrometry.* Nat Biotechnol, 1999. **17**(7): p. 676-82.

19. Yuvaniyama, J., et al., *Crystal structure of the dual specificity protein phosphatase VHR.* Science, 1996. **272**(5266): p. 1328-31.

20. Yokota, T., et al., *Crystal structure of human dual specificity phosphatase, JNK stimulatory phosphatase-1, at 1.5 A resolution.* Proteins, 2007. **66**(2): p. 272-8.

21. Agarwal, R., S.K. Burley, and S. Swaminathan, *Structure of human dual specificity protein phosphatase 23, VHZ, enzyme-substrate/product complex.* J Biol Chem, 2008. **283**(14): p. 8946-53.

22. Rinner, O., et al., *Identification of cross-linked peptides from large sequence databases.* Nat Methods, 2008. **5**(4): p. 315-8.

23. Seebacher, J., et al., *Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing.* J Proteome Res, 2006. **5**(9): p. 2270-82.

APPENDIX

Press license for LEA protein publication

Sharon, M.A., Kozarova, A., Clegg, J.S., Vacratsis, P.O., and Warner, A.H. (2009). Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp Artemia franciscana. Biochem Cell Biol *87*, 415-430

NRC RESEARCH PRESS LICENSE
TERMS AND CONDITIONS

Apr 27, 2010

This is a License Agreement between Anna Kozarova ("You") and NRC Research Press ("NRC Research Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by NRC Research Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2417261502723
License date	Apr 27, 2010
Licensed content publisher	NRC Research Press
Licensed content publication	Biochemistry and Cell Biology
Licensed content title	Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp *Artemia franciscana*
Licensed content author	Michelle A. Sharon, Anna Kozarova, James S. Clegg, et al
Licensed content date	Apr 1, 2009
Volume number	87
Issue number	2
Type of Use	Thesis/Dissertation
Requestor type	Author (original work)
Format	Electronic
Portion	Full article
Order reference number	
Title of your thesis / dissertation	Characterization of cellular systems using biological mass spectrometry
Expected completion date	May 2010
Estimated size(pages)	130
Total	0.00 USD
Terms and Conditions	

General Terms & Conditions

Permission is granted upon the requester's compliance with the following terms and conditions:

1. A credit line will be prominently placed in your product(s) and include: for books the author, book title, editor, copyright holder, year of publication; for journals the author, title of article, title of journal, volume number, issue number, and the inclusive pages. The credit line must include the following wording: "© 2008 NRC Canada or its licensors. Reproduced with permission," except when an author of an original article published in 2009 or later is reproducing his/her own work.

2. The requester warrants that the material shall not be used in any manner that may be derogatory to the title, content, or authors of the material or to National Research Council Canada, including but not limited to an association with conduct that is fraudulent or otherwise illegal.

3. Permission is granted for the term (for Books/CDs-Shelf Life; for Internet/Intranet-In perpetuity; for all other forms of print-the life of the title) and purpose specified in your request. Once term has expired, permission to renew must be made in writing.

4. Permission granted is nonexclusive, and is valid throughout the world in English and the languages specified in your original request. A new permission must be requested for revisions of the publication under current consideration.

5. National Research Council Canada cannot supply the requester with the original artwork or a "clean copy."

6. If the National Research Council Canada material is to be translated, the following lines must be included: The authors, editors, and National Research Council Canada are not responsible for errors or omissions in translations.

v1.3

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK10775117.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2417261502723 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Press license for ICAT publication

Kozarova, A., Sliskovic, I., Mutus, B., Simon, E.S., Andrews, P.C., and Vacratsis, P.O. (2007). Identification of redox sensitive thiols of protein disulfide isomerase using isotope coded affinity technology and mass spectrometry. J Am Soc Mass Spectrom *18*, 260-269

ELSEVIER LICENSE
TERMS AND CONDITIONS

Apr 27, 2010

This is a License Agreement between Anna Kozarova ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Anna Kozarova
Customer address	Department of Chemistry and Biochemistry Windsor, ON N9B 3P4
License Number	2417270439001
License date	Apr 27, 2010
Licensed content publisher	Elsevier
Licensed content publication	Journal of the American Society for Mass Spectrometry
Licensed content title	Identification of Redox Sensitive Thiols of Protein Disulfide Isomerase Using Isotope Coded Affinity Technology and Mass Spectrometry
Licensed content author	Anna Kozarova, Inga Sliskovic, Bulent Mutus, Eric S. Simon, Philip C. Andrews, Panayiotis O. Vacratsis
Licensed content date	February 2007
Volume number	18
Issue number	2
Pages	10
Type of Use	Thesis / Dissertation
Portion	Full article
Format	Both print and electronic
You are an author of the Elsevier article	Yes
Are you translating?	No
Order Reference Number	

Expected publication date	May 2010
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD
Total	0.00 USD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and
he permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx , As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com

All content posted to the web site must maintain the copyright information line on the bottom of each image
You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the

material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**: None

v1.6

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK10775121.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2417270439001 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

VITA AUCTORIS

Anna Kozarova

Education: University of Windsor, Windsor, Ontario
2005-2010 (Ph.D.)

Laurentian University, Sudbury, Ontario
2002-2004 (M.Sc.)

McMaster University, Hamilton, Ontario
1997-2000 (B.Sc.)

Employment: Samuel Lunenfeld Research Institute, Toronto, Ontario
2000-2002