

Université de Montréal

**Large Scale Identification of Protein SUMOylation by  
Mass Spectrometry in HEK293 Cells**

par

Louiza Mahrouche

Département de Biochimie, Institut de recherche en immunologie et oncologie  
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures et postdoctorales  
en vue de l'obtention du grade de Maîtrise ès science (M.Sc.)  
en Biochimie

décembre 2010

©Louiza Mahrouche, 2010

Université de Montréal  
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Large Scale Identification of Protein SUMOylation by Mass Spectrometry in HEK293  
cells

Présentée par :

Louiza Mahrouche

a été évalué par un jury composé des personnes suivantes :

Dr. Muriel Aubry, président-rapporteur  
Dr. Pierre Thibault, directeur de recherche  
Dr. Benoît Coulombe, membre du jury

## Résumé

Une large gamme d'événements cellulaires est régulée par la SUMOylation des protéines. Cette modification post-traductionnelle est impliquée dans le cancer notamment dans la leucémie promyélocytaire aigue. À ce jour, peu d'études à grande échelle ont porté sur l'identification des sites de modification. Ce mémoire présente une approche protéomique quantitative unique qui combine une double purification par affinité au niveau des protéines cibles ainsi que des peptides modifiés.

L'approche la plus répandue de purification des protéines SUMOylés implique l'utilisation d'une forme de SUMO modifié avec une étiquette (His<sub>6</sub>-SUMO). A ce jour, les approches permettant l'enrichissement au niveau peptidique nécessite une forme mutante de SUMO.

Notre analyse consiste à premièrement enrichir en protéines SUMOylés dans les cellules humaines vierges ou sur exprimant His<sub>6</sub>-SUMO-1/3 en présence ou pas de trioxyde de diarsenic, un traitement de leucémie promyélocytaire aigue. Par la suite, les échantillons sont digérés et les peptides obtenus des protéines SUMOylés conservent un branchement caractéristique. Les peptides sont soit immunoprécipités avec un anticorps spécifique au branchement SUMO ou directement analysés par nano LC/LC-MS/MS par un spectromètre de masse LTQ-Orbitrap. Une analyse manuelle des données révèle des fragments caractéristiques correspondant à la chaîne latérale de SUMO. L'originalité de l'approche réside dans l'identification quantitative et sans ambiguïté des sites de SUMOylation. Cette approche a permis l'identification de 17 et 3 sites de SUMO-3 et SUMO-1 respectivement dans les cellules HEK293. Finalement, la SUMOylation de PML est induite suite au traitement d'arsenic.

**Mots-clés :** SUMOylation, spectrométrie de masse, purification par affinité, identification de sites de SUMOylation

## Abstract

A wide range of cellular events are regulated by protein SUMOylation. This posttranslational modification was involved in APL (acute promyelocytic leukemia). Only a few large scale studies in mammalian cells have focused on identifying the conjugation sites. This thesis presents a unique quantitative proteomics approach that combines double affinity purification at the protein and peptide level.

A common approach to purification of SUMOylated proteins involves the use of a tagged SUMO (His<sub>6</sub>-SUMO). To date, the SUMO peptide isolation is addressed using an engineered SUMO.

In presence or absence of arsenic trioxide, a treatment of APL, mock and His<sub>6</sub>-SUMO1/3 expressing cells are lysed and the SUMOylated proteins are isolated under denaturing conditions. Subsequently, these samples are digested and the peptides bearing the modification site bear a specific SUMO stub. They are either immunoprecipitated with an anti SUMO stub antibody or directly analyzed by nano LC coupled to an LTQ-Orbitrap mass spectrometer. Manual analysis of the data reveals characteristic fragmentation corresponding to the side chain of SUMO. The originality of the approach lies in the quantitative and unambiguous identification of SUMOylation sites *in vivo*. This approach allowed the identification of 17 and 3 sites of SUMO-3 and SUMO-1, respectively, in HEK293 cells. Finally, PML was identified as the major SUMOylation target following arsenic treatment.

**Keywords:** SUMOylation, mass spectrometry, affinity purification, identification of SUMOylation sites.

## Table of Contents

RÉSUMÉ.....	I
ABSTRACT .....	II
TABLE OF CONTENTS .....	III
LIST OF TABLES .....	VI
LIST OF FIGURES.....	VII
REMERCIEMENTS .....	X
ABBREVIATIONS.....	XI
1. INTRODUCTION.....	1
1.1 SUMOYLATION PROCESS IN THE CELL .....	2
1.1.1 Discovery of SUMO.....	2
1.1.2 SUMO conjugation, processing and deconjugation.....	3
1.1.2.1 SUMO conjugation.....	4
1.1.2.2 SUMO deconjugation .....	6
1.1.2.2.1 Endopeptidase activity.....	6
1.1.2.2.2 Isopeptidase activity .....	7
1.1.2.3 Poly-SUMO chain formation.....	7
1.1.3 Consensus sequence .....	8
1.1.4 Molecular consequences of SUMOylation .....	9
1.2 MASS SPECTROMETRY .....	10
1.2.1 Ionization.....	11
1.2.2 The Mass Analyzer.....	13
1.2.2.1 The Orbitrap mass analyzer .....	14
1.2.2.2 Linear ion trap mass analyzer .....	15
1.2.3 Tandem MS .....	15
1.2.4 Detectors.....	18

1.2.5	Database Searching .....	18
1.2.6	Quantitative Proteomics .....	22
1.3	SUMO MS BASED PROTEOMICS .....	24
1.3.1	Enrichment of SUMOylated proteins.....	24
1.3.2	Enrichment strategies of SUMO-modified peptides .....	25
1.3.3	Stress-induced SUMOylation.....	27
1.3.4	Spectral interpretation strategies .....	27
1.4	PROJECT'S GOAL .....	31
1.5	REFERENCES .....	33
2. A NOVEL PROTEOMICS APPROACH TO IDENTIFY SUMOYLATED PROTEINS AND THEIR MODIFICATION SITES IN HUMAN CELLS.....		41
2.1	ABSTRACT .....	42
2.2	INTRODUCTION.....	43
2.3	EXPERIMENTAL SECTION .....	47
2.4	RESULTS.....	53
2.5	DISCUSSION.....	64
2.6	ACKNOWLEDGEMENTS .....	69
2.7	REFERENCES .....	81
3. DISCUSSION .....		88
3.1	GENERAL APPROACH .....	89
3.2	<i>In Vitro</i> SUMOYLATION.....	89
3.3	THE HIS-SUMO MUTANT FUNCTIONALITY IN HEK293 CELLS.....	92
3.4	OVEREXPRESSION OF HIS-SUMO PROTEINS IN HEK293 CELLS.....	93
3.5	SUBCELLULAR DISTRIBUTION OF SUMO .....	95
3.6	Ni-NTA ENRICHMENT .....	96
3.7	DIGESTION AND PEPTIDE SEPARATION.....	97
3.8	SUMO PEPTIDE IDENTIFICATION BY MASS SPECTROMETRY .....	99
3.8.1	Sample Complexity .....	99

3.8.2	Spectral interpretation strategies .....	99
3.8.3	Fragmentation methods .....	102
3.9	IMMUNOPRECIPITATION AND SUMO-1 SITE IDENTIFICATION .....	102
3.10	EFFECT OF ARSENIC TRIOXIDE .....	104
3.11	IDENTIFIED TARGETS .....	106
3.12	REFERENCES .....	109
	4. CONCLUSION .....	116
4.1	THE SUMO CHALLENGE .....	117
4.2	FUTURE PERSPECTIVES.....	118
4.3	REFERENCES .....	119
	ANNEX I: SUPPLEMENTARY FIGURES 2.1 -2.9 FOR CHAPTER 2 .....	I
	ANNEX II: SUPPLEMENTARY FIGURE 2.10 FOR CHAPTER 2 .....	XIX
	ANNEX III: SUPPLEMENTARY FIGURE 2.11 FOR CHAPTER 2 .....	XLVI
	ANNEX IV: SUPPLEMENTARY TABLES I-III FOR CHAPTER 2.....	LI

## List of Tables

Table I : List of confirmed SUMOylated peptides from NTA-enriched of nuclear proteins from HEK293 His <sub>6</sub> -SUMO-3 mutant .....	70
Table II :List of confirmed SUMOylated peptides from the immunoprecipitation of nuclear proteins from HEK293 His6-SUMO-1 mutant.....	71
Supplementary Table I: Peptide and protein identification from NTA affinity-purified mock HEK293 and HEK293 His <sub>6</sub> -SUMO3 mutant nuclear extracts with and without As <sub>2</sub> O <sub>3</sub> . Table on CD ROM.....	lii
Supplementary Table II: Peptide and protein identification from dual affinity purification (NTA and immunoprecipitation) of mock HEK293 and HEK293 His <sub>6</sub> -SUMO1 mutant nuclear extracts with and without As <sub>2</sub> O <sub>3</sub> . Table on CD ROM.....	lii
Supplementary Table III : Identification of SUMOylated peptides with Mascot, ChopNSpice and SUMmOn from NTA affinity-purified HEK293 His <sub>6</sub> -SUMO3 mutant nuclear extracts with and without As <sub>2</sub> O <sub>3</sub> . .....	liii

## List of Figures

Figure 1.1: Overlay SUMO-1 (black) and ubiquitin (red) .....	3
Figure 1.2: SUMO pathway: activation, conjugation and deconjugation.....	4
Figure 1.3: Isopeptide link between SUMO and its target protein .....	5
Figure 1.4: Reactions catalyzed by SENPs.....	6
Figure 1.5: Use of ESI in an LC-MS coupled instrument.....	12
Figure 1.6: Peptides are ionized in the ESI and fragmented in either the LTQ or in the HCD collision cell. The Orbitrap is used to detect parent m/z.....	14
Figure 1.7: Nomenclature of fragment ions observed for the dissociation of peptide ions .	17
Figure 1.8: The MS/MS spectrum is compared against a theoretical MS/MS spectrum generated <i>in silico</i> . A score is given to each peptide based on the similarity of the two spectra .....	21
Figure 1.9: Contour map created by Proteoprofile.....	23
Figure 1.10: Purification strategy of His-SUMO conjugated proteins using the Ni-NTA pull down.....	24
Figure 1.11: Theoretical MS/MS CID spectra of a) unmodified peptide, b) phospho-peptide and c) SUMO-modified peptide.....	29
Figure 2.1: Overview of proteomics approach to the identification of SUMOylated peptides. ....	72
Figure 2.2: In vitro SUMOylation assays .....	73

Figure 2.3 : Comparison of His-SUMO WT and mutants to SUMOylate PML and to colocalize within nuclear bodies .....	74
Figure 2.4: Immunoblots of NTA purified extracts from control HEK293 and HEK293-His-SUMO cells.....	75
Figure 2.5: Mass spectrometry analyses of SUMOylated proteins from extracts of control HEK293 and HEK293 His <sub>6</sub> -SUMO-3 in As <sub>2</sub> O <sub>3</sub> -stimulated and non-stimulated cell..	77
Figure 2.6: Identification of new SUMOylation sites in the protein promyelocytic leukemia (PML).....	79
Figure 2.7: Identification of proteins modified by SUMO-1 using a combined NTA/immunoaffinity enrichment approach.....	80
Supplementary Figure 2.1: MS analyses of <i>in vitro</i> digestion products of GST-RanGAP1 (418-587) confirmed the SUMOylation site K524 for all SUMO paralogs. ....	iii
Supplementary Figure 2.2: MS analyses of <i>in vitro</i> SUMOylation enabled the identification of sites of polySUMOylation. ....	vii
Supplementary Figure 2.3 : <i>in vitro</i> deSUMOylation assay of E2-ligase and RanGAP1 using SENP1 .....	ix
Supplementary Figure 2.4 : Comparison of His-SUMO1 WT and His-SUMO1 mutant to SUMOylate PML and to colocalize within nuclear bodies.....	x
Supplementary Figure 2.5: Analysis of Gene Ontology terms associated with biological processes .....	xi
Supplementary Figure 2.6: NTA purified extracts from control HEK293 and HEK293-His-SUMO cells.....	xiii

Supplementary Figure 2.7: Tandem mass spectra of synthetic peptides to confirm SUMOylation of sites K380 and K400 .....	xiv
Supplementary Figure 2.8: Design of an immunoprecipitation experiment using a centrifugal filter device and Recovery yields of SUMOylated peptides.....	xvi
Supplementary Figure 2.9: Mass spectrometry analyses of SUMOylated proteins from Immunoaffinity purified NTA extracts of mock HEK293 and HEK293 His <sub>6</sub> -SUMO1 in As <sub>2</sub> O <sub>3</sub> -stimulated and non-stimulated cells.....	xviii
Supplementary Figure 2.10: Tandem Mass Spectra of All Identified SUMO-3 Peptides Following Database Search and Manual Validation.....	xix
Supplementary Figure 2.11: Tandem Mass Spectra of All Identified SUMO-1 Peptides Following Database Search and Manual Validation.....	xlvi

## Remerciements

D'abord, je remercie mon directeur de recherche le Dr Pierre Thibault pour son accueil ainsi que les connaissances scientifiques qu'il m'a transmises. Son grand intérêt ainsi que la confiance qu'il m'a témoignée ont permis à ce projet de progresser

Je souhaite également remercier vivement Frédéric Galisson pour sa très grande contribution au projet SUMO. Son excellent travail, son esprit critique et sa grande expérience ont permis au projet SUMO de voir le jour mais aussi de se développer. Merci à Dr Mounira Chelbi-Alix pour sa collaboration et son accueil.

Je désire remercier le Dr Eric Bonneil pour son expertise sur la spectrométrie de masse et le partage de ses connaissances, Mathieu Courcelles pour son assistance aux outils bioinformatiques ainsi que Christelle Pomiès pour son aide dans les différentes techniques de biologie. De plus, je remercie sincèrement Chantal Durette, Robert Gomma et Jean-Michel Laprise pour leur aide lors de l'écriture de ce mémoire. Merci Chantal de prendre le relais du projet. Mes remerciements s'adressent également à tout mes collègues de laboratoire notamment Gaëlle Bridon, Marlene Gharib et Tara Muratore pour avoir créé un environnement d'entraide et amical.

Un grand merci à ma sœur Lilia et à Nicolas ainsi que tous mes proches qui m'ont soutenue durant ces deux dernières années.

Pour terminer, je remercie le FQRNT ainsi que la Faculté des études supérieures de l'Université de Montréal pour leur soutien financier.

## Abbreviations

ACN	Acetonitrile
ADP	Adenosine diphosphate
APL	Acute Promyelocytic Leukemia
As <sub>2</sub> O <sub>3</sub>	Arsenic trioxide
ATP	Adenosine-5'-triphosphate
CI	Chemical ionization
CID	Collision induced dissociation
Daxx	Death domain-associated protein
DMEM	Dulbecco's modified eagle medium
E2-25K	E2 ubiquitin ligase – 25 kilo dalton
EDTA	Ethylenediaminetetraacetic acid
EI	Electron ionization
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FPR	False positive rate
FTMS	Fourier transform mass spectrometer
GST	Glutathione S-transferase
HCD	Higher-energy C-trap dissociation
HEK	Human embryony kidney
hnRNP	Heterogeneous nuclear ribonucleoprotein
HPLC	High-performance liquid chromatograph
IMAC	Immobilized metal affinity chromatography
IP	Immunoprecipitation
kDa	Kilo dalton
LC	Liquid chromatography
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LIT	Linear ion trap

LTQ	Linear trap quadrupole
m/z	Mass/charge
MALDI	Matrix-assisted laser desorption ionization
Mdm2	Murine double minute-2
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MudPit	Multidimensional protein identification technology
MW	Molecular weight
NB	Nuclear body
NDSM	Negatively charged amino acid-dependent SUMOylation motif
NEM	N-Ethylmaleimide
NFkB	Nuclear factor kB
Ni-NTA	Nickel-nitriloacetic acid
NLS	Nuclear localization signal
PARP1	Poly (ADP-ribose) polymerase 1
PCNA	Proliferating cell nuclear antigen
PDSM	Phosphorylation-dependent sumoylation motif
PIAS	Protein inhibitor of activated STAT
PML	Promyelocytic leukemia protein
ppm	Parts per million
PTM	Post-translational modification
RanBP2	Ran Binding Protein 2
RanGap	Ran GTPase activating protein
RAR- $\alpha$	Retinoic acid receptor alpha
RBCC	Ring finger B-box coiled-coil
RING	Really interesting new gene
RNF4	RING finger 4 protein
RSF1	Remodeling and spacing factor 1
SAE	SUMO activating enzyme

SAFB2	Scaffold attachment factor B 2
SENP	SUMO-specific isopeptidase
SIM	SUMO interacting motif
STAT	Signal transducer and activator of transcription protein
SUMO	Small ubiquitin-like modifier
TAP	Tandem Affinity Purification
TBS	Tris buffered saline
TIF1b (TRIM28)	Transcription intermediary factor 1-beta (TRIM28)
Ub	Ubiquitin
Ubc9	Ubiquitin-like protein SUMO-1 conjugating enzyme
Ubl	Ubiquitin-like protein
WT	Wild type
YFP	Yellow fluorescent protein

# **1. Introduction**

## 1.1 SUMOylation process in the cell

Post-translational modifications (PTMs) represent universal and fundamental mechanisms by which protein function, activity, stability and localization can be regulated. These modifications extend significantly the diversity of the proteome. PTMs are divided into two groups: proteolytic cleavages and covalent modifications. There are over 150 different covalent modifications on a variety of amino acid side chains (Voet and Voet 1995), among these are small chemical group modifications such as acetylation, phosphorylation or larger macromolecules attachment such as ubiquitylation and SUMOylation, the latter being the primary focus of the present study.

### 1.1.1 Discovery of SUMO

In the 1970's, the first protein acting as a ubiquitin-like (UBL) modifier was discovered : ubiquitin (Hochstrasser 2009). However, the first documented report of a related small ubiquitin modifier (SUMO) was only made 20 years later by Mahajan *et al.* for Ran GTPase 1 (RanGap1) covalently modified by SUMO-1 in mammals (Mahajan, Delphin et al. 1997). At the present day, at least nine UBLs are shown to covalently modify their targets, and it is suspected that additional UBLs are likely to be discovered in the future (Hochstrasser 2009). Regardless of their low sequence similarity to ubiquitin, all UBLs share a common 3D structure: the ubiquitin fold. For instance, although SUMO-1 and ubiquitin share only 18% sequence identity, SUMO has the characteristic  $\beta\beta\alpha\beta\beta\alpha\beta$  fold of the ubiquitin protein family (Bayer, Arndt et al. 1998) as observed in Figure 1.1.

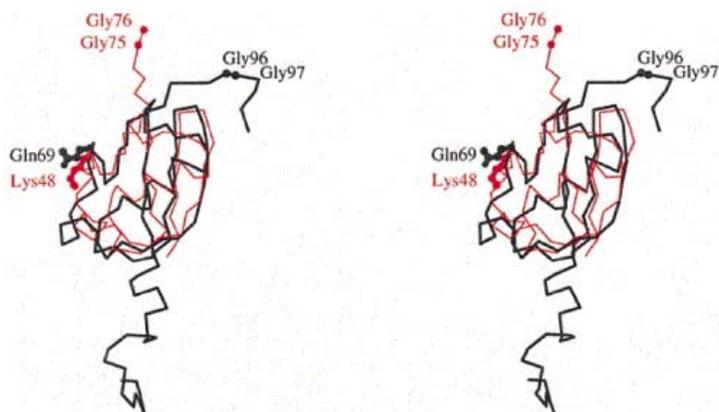


Figure 1.1: Overlay SUMO-1 (black) and ubiquitin (red) (Bayer, Arndt et al. 1998)

SUMO is expressed in all eukaryotic cells and in different cell types in multicellular organisms. Yeasts have a single isoform of SUMO, while in vertebrates, four paralogs of SUMO (SUMO1-4) are expressed (Saitoh and Hinchev 2000), and in plants eight versions of SUMO have been identified (Kurepa, Walker et al. 2003).

In humans, SUMO-1 is the most studied paralog and hundreds of SUMO1-2-3 conjugates have been identified. SUMO1 and SUMO2/3 share about 50% similarity, but SUMO-2 and SUMO-3 differ only by 4 residues. SUMO1 and SUMO2/3 seem to have different functions since they conjugate different substrates *in vivo* and have different responses to stress (Saitoh and Hinchev 2000). The role of SUMO-4 is still unknown, no conjugates have been detected *in vivo* and its *in vivo* maturation into a conjugation-competent form still remains unclear (Bohren, Nadkarni et al. 2004).

### 1.1.2 SUMO conjugation, processing and deconjugation

UBLs not only share a common structure, but they use similar conjugation mechanisms requiring a multi-stage ATP-dependent enzymatic cascade (Hochstrasser 2009). Specific and highly dynamic machinery is responsible for SUMO conjugation: the formation of an isopeptide bond between the C-terminal glycine of the mature SUMO and the amino  $\epsilon$  group of the target's specific lysine residue. The specificity of



SUMO protein is transferred from Sae2 to UBC9 and forms a thioester link with cys93 of UBC9. Note that UBC9 contains a SIM (SUMO-interacting motif) domain capable of recognizing the SUMO protein. Also UBC9-SUMO complex is capable of substrate specificity through the consensus SUMOylation motif found on the target protein (Tatham, Jaffray et al. 2001). At this point, UBC9 can either directly transfer SUMO protein to its substrate or optionally *in vivo* this step may require the cooperation of an E3 ligase.

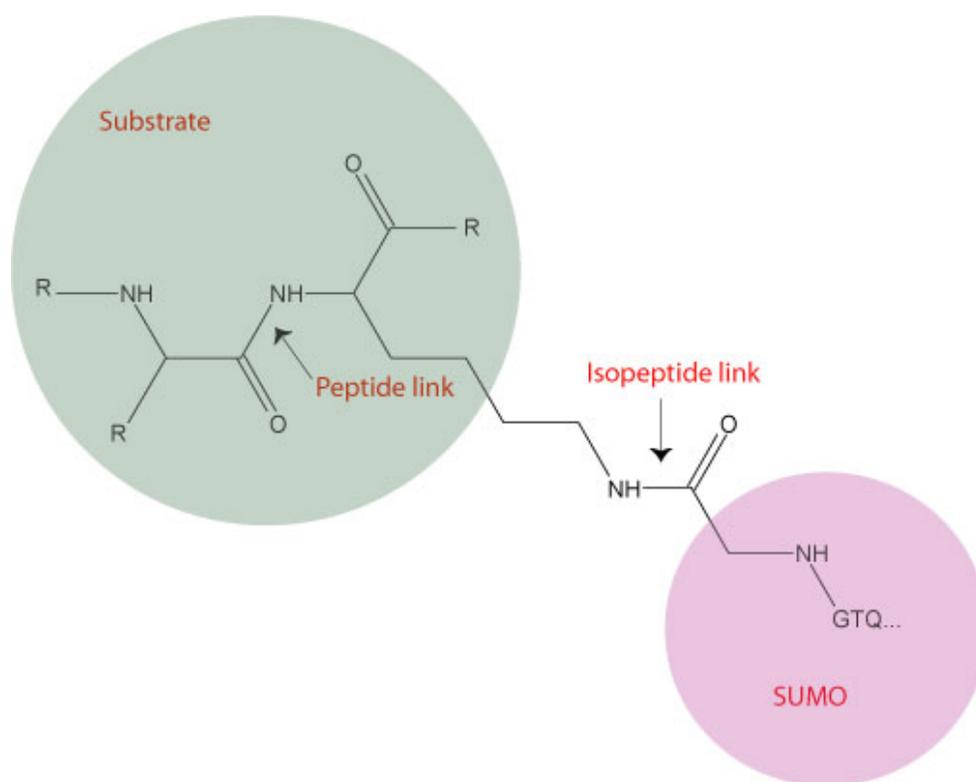


Figure 1.3: Isopeptide link between SUMO and its target protein

A certain number of E3 ligases have been identified for the SUMO pathway. They are believed to facilitate the formation of the isopeptide bond between the C-terminal SUMO protein and the acceptor's lysine by forming a complex with UBC9 and the SUMO protein. E3 enzymes do not seem to form thioester linkage with SUMO, but function as scaffold proteins bringing UBC9-SUMO complex in close contact with the substrate. E3 ligases have two roles: i) enhance the SUMO conjugation, and ii)

participate in the specificity of the substrate. Most of the identified E3 ligases are nuclear, although new studies have reported the existence of a mitochondrial SUMO E3 (Braschi, Zunino et al. 2009). Examples of E3 ligases include PIAS family, HDAC4 and RanBP2 (Wilkinson and Henley 2010).

### 1.1.2.2 SUMO deconjugation

SUMO is removed from its substrate by a family of enzymes known as the sentrin-specific proteases (SENPs). Seven SUMO-specific proteases have been identified in the human genome (Marcin Drag 2008), showing different specificities towards SUMO isoforms. SENPs are cysteine proteases and have two main enzymatic activities: endopeptidase and isopeptidase as described in Figure 1.4.

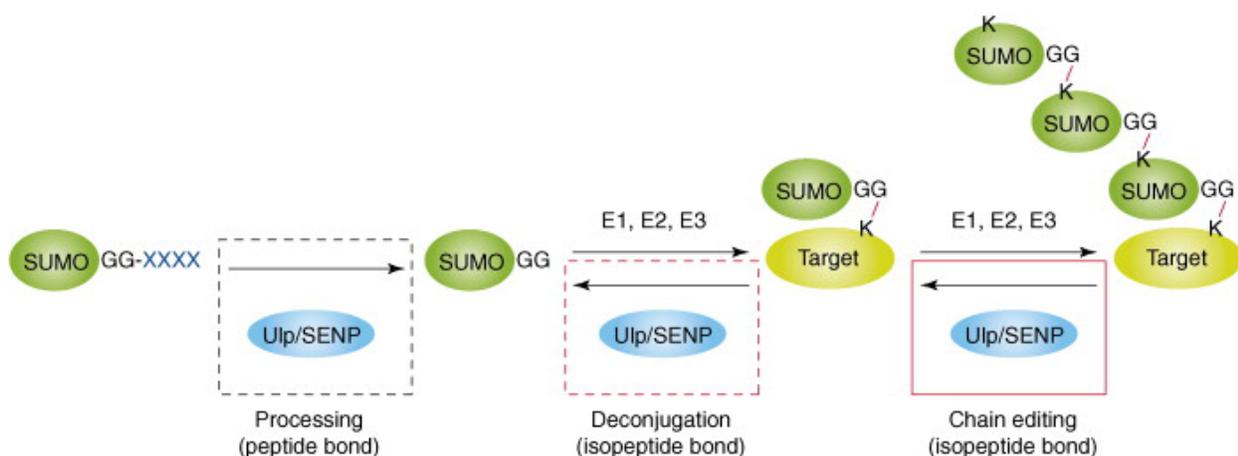


Figure 1.4: Reactions catalyzed by SENPs. Adapted (Mukhopadhyay and Dasso 2007)

#### 1.1.2.2.1 Endopeptidase activity

Newly synthesized SUMO is in an immature form and prior to conjugation, it needs to be activated by SENPs (see Figure 1.2). The propeptide is cleaved by SENP to reveal the essential diglycine motif which is conserved among all ubiquitin-like proteins

(Mukhopadhyay and Dasso 2007). The free diglycine motif is essential for the conjugation reaction.

#### **1.1.2.2 Isopeptidase activity**

The second role of SENP is the cleavage of the isopeptide bond (Figure 1.3), which is formed between the target and the SUMO protein. Therefore not only do SENPs control the pool of available SUMO, but they are also responsible for the half life of the conjugated species (Mukhopadhyay and Dasso 2007). Moreover, SENPs also control the poly-SUMO chain formation, by hydrolyzing the SUMO-SUMO bond. Due to their central roles, SENPs are believed to be highly regulated.

In summary the SUMO pathway is highly regulated, first by its conjugation using a cascade of enzymes comprising an E1, an E2 and multiple E3 enzymes, but also by its deconjugation pathway.

#### **1.1.2.3 Poly-SUMO chain formation**

The concept of a modifier being modified adds a new layer of complexity, but also provides versatility to protein function. The formation of poly-ubiquitin chains has been extensively studied and it became clear that a specific poly-Ub chain structure provides a specific biological outcome (Pickart and Fushman 2004). The wide array of ubiquitylation outcome can be explained by the diversity of ubiquitin chain structure: the chain length but most importantly the different cross linkages that could be formed.

Poly-SUMO chain formations have not been extensively studied due to important technical limitations: the half life and low abundance of the SUMO conjugates, as well as the absence of fast and straightforward method for the identification of SUMOylation sites. So far, SUMO2/3 has been shown to form poly-SUMOylation sites *in vivo* (Tatham, Jaffray et al. 2001), through the internal K11 that lies in a consensus sequence that is missing in SUMO-1 (Matic, Van Hagen et al. 2008). Although SUMO-1 is also

incorporated into the SUMO2/3 chains, it can't be further modified and seems to work as a capping modifier. However, *in vitro* SUMO-1 has been shown to form polymers through K7,16,17,37,39,46 (Cooper, Tatham et al. 2005; Pedrioli, Raught et al. 2006). It is important to note that *in vitro* conditions are artificial (high concentrations of E1 and E2, absence of SENPs or E3 ligase) and can introduce artifacts.

### 1.1.3 Consensus sequence

Through the use of known SUMOylation sites of RanGap1, PML, p53 and SP100 a SUMOylation acceptor motif has been identified as  $\psi$ -K-X-E/D (Rodriguez, Dargemont et al. 2001) (where K is the acceptor lysine and  $\psi$  is a hydrophobic residue). No consensus sequence has been identified for ubiquitylation that might be due in part to the presence of a single E2 enzyme for SUMO (namely UBC9) that recognizes the consensus sequence and can directly SUMOylate the target. In contrast, the ubiquitin pathway contains about 20 E2 ligases and hundreds of E3 ligases (Semple 2003).

Although 75% of known SUMO substrates are modified at the  $\psi$ -K-X-E/D motif (Xue, Zhou et al. 2006), this proportion is probably overestimated. Recently, Matic *et al* have proposed a new consensus motif: the inverted motif (E/DxK) (Matic, Schimmel et al. 2010) that seems to be less common.

Two extended motifs have been discovered PDSM and NDSM. PDSM is a SUMOylation motif ( $\psi$ -K-X-E/D-X-X-pS-P) where downstream phosphorylation enhances SUMOylation (Hietakangas, Anckar et al. 2006). NDSM is characterized by the presence of the core consensus sequence followed by negatively charged residues at the C-terminus of the acceptor lysine (Yang, Galanis et al. 2006). A hydrophobic cluster SUMOylation motif was identified on 16 sites and its presence seems to greatly increase the efficiency of SUMOylation (Matic, Schimmel et al. 2010). The presence of a consensus is not indicative that the protein is modified. As revealed by crystal structure studies, the SUMOylation consensus sequence has been shown to be recognized

specifically by UBC9 when found in an unstructured region of a protein (Bernier-Villamor, Sampson et al. 2002).

Since 25% of known SUMOylation sites are non-consensus sites, the mechanism through which these substrates are recognized by UBC9 is still unknown.

Because of the difficulty in identifying the SUMOylation site on a given potential substrate, *in silico* prediction tools that make use of consensus motifs have emerged such as SUMOsp (Xue, Zhou et al. 2006).

#### **1.1.4 Molecular consequences of SUMOylation**

Protein SUMOylation is associated with numerous cell functions. In contrast to ubiquitylation that is best known to target protein for degradation, it is not possible to predict the biological outcome of SUMOylation on a given target. As for other PTMs, SUMO has been shown to be implicated in diverse and multiple biological mechanisms: intracellular transport, regulation of transcription and protein degradation. In the present section, I will cover one of the most studied functions of SUMO: Promyelocytic Leukaemia Protein Nuclear Bodies (PML-NBs) regulation.

Nuclear bodies (NBs) are discrete protein aggregates where PML functions as the main scaffold protein. NBs have been implicated in multiple cellular functions such as transcriptional regulation and apoptosis, they are highly dynamic and a large number of SUMOylated proteins lie within this structure (Van Damme, Laukens et al. 2010). Because they are associated with a high number of cellular disorders (Lallemand-Breitenbach and de Thé 2010), multiple studies in the past years have been conducted in understanding the composition as well as the function of NBs. More than 50 different proteins have been shown to shuffle in and out of NBs (Reineke, Liu et al. 2009), and in order to gain a better understanding of these structures, great efforts were made to characterize the complex regulation of PML. NBs were proposed to act as SUMOylation platforms, since a high proportion of NBs' proteins are either SUMOylated or contain a

SIM domain. Moreover, the SIM-SUMO interaction might account for protein recruitment into the NBs (Lallemand-Breitenbach and de The 2010).

PML is essential to nuclear body formation and seems to be the main recruiter of all the components, although it mainly interacts indirectly with NB proteins. PML is highly regulated at the transcriptional level but also by post-translational modifications such as SUMOylation. PML has been shown to be modified by all SUMO paralogs at K65,160 and 490 (Kamitani, Nguyen et al. 1998; Ayaydin and Dasso 2004) and it contains a SIM domain. It has been shown that SUMOylation of PML promotes its subsequent ubiquitylation and subsequent degradation, and this process is enhanced by arsenic trioxide ( $As_2O_3$ ), a drug that is used in treating APL (Lallemand-Breitenbach, Jeanne et al. 2008).

## 1.2 Mass Spectrometry

Mass spectrometry is an analytical technique used to determine the elemental composition, structural information as well as the amount of analyte. The mass spectrometer separates and measures gas-phase ion's  $m/z$  (mass to charge ratio). Mass spectrometry qualitative and quantitative applications are diverse and numerous and the focus in this thesis will be on MS based proteomics.

The mass spectrometer can only detect molecules that are ionized; therefore the first step is the ionization and the vaporization of the molecule. Once the ion is in the gas phase it is separated by the mass analyzer based on its  $m/z$ . The observed ion can then be fragmented in order to obtain structural information. Finally, the abundance of the separated ion is recorded by the detector.

In a typical large scale proteomics experiment, the analyte is the peptide mixture. The general procedure usually starts with cells lysis and can be followed by sub-proteome fractionation. The proteins of choice are then enzymatically digested into peptides, usually by trypsin. The peptides are then chromatographically separated based

on their charge and hydrophobicity and analyzed by MS/MS. The raw data (MS and MS/MS spectra) are submitted to database search engines and a list of peptides along with their PTMs is obtained.

### **1.2.1 Ionization**

Mass spectrometry (MS) has long been used mainly to analyze small volatile organic molecules, mostly because of the limitations of the classical ionization systems: electron ionization (EI) and chemical ionization (CI). With the introduction of electrospray ionization, ESI (Fenn, Mann et al. 1989), and matrix-assisted laser desorption ionization, MALDI (Karas and Hillenkamp 1988), the mass spectrum of intact peptides and proteins could be obtained. ESI and MALDI both present two main characteristics that make it possible to analyze biomolecules: i) soft ionization that does not disrupt the molecule, and ii) vaporization of non-volatile compounds.

One of ESI's main advantages is the possibility of using MS coupled to a liquid chromatography (LC) system (Figure 1.5). The LC-MS combination represented two main challenges. First the removal of all the solvent coming from the LC, because the MS operates under vacuum, and second, the production of gas-phase ions from the normally non-volatile LC's analytes (Watson and Sparkman 2007). Thus, ESI became the most successful interface for LC/MS, as the majority of the solvent is evaporated.

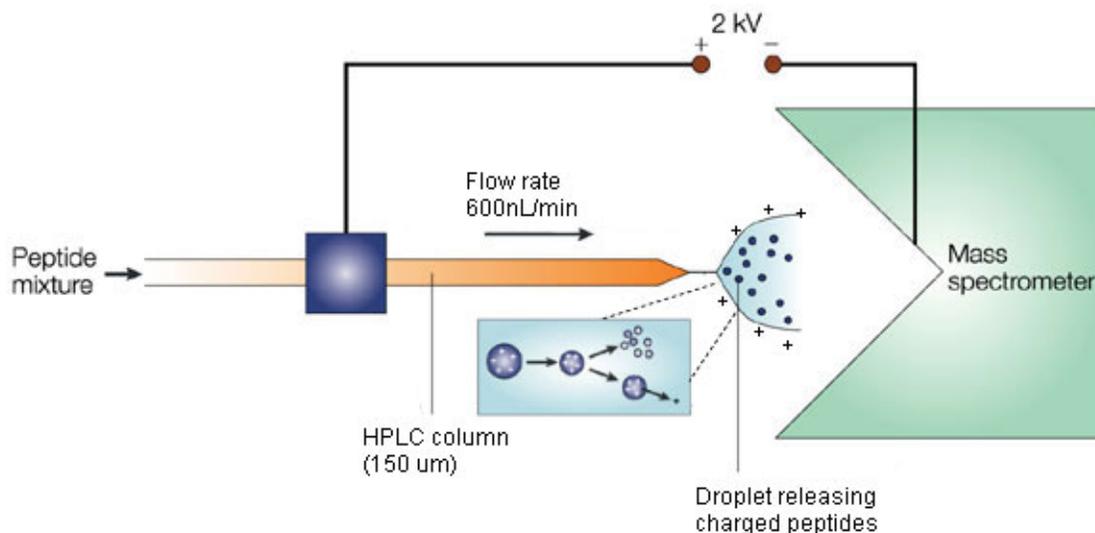


Figure 1.5: Use of ESI in an LC-MS coupled instrument (adapted from (Steen and Mann 2004))

In ESI the droplets containing the analyte are formed after the solution has been forced through a very small capillary (20 μm internal diameter) in the presence of an electric field. The electric field is fundamental for the ionization, but also for the nebulization of the analyte solution into fine droplets. As the volatile solvent (water and acetonitrile mixture) evaporates from the droplets, the ratio of charge to droplet size increases leading to charge repulsion. This process ends with a coulombic explosion, where smaller droplets are produced and analyte ions are ejected. The droplets eventually desolvate completely after multiple cycles of coulombic explosions. ESI can be used in positive and negative modes, by switching the potential, but it is mostly used in positive mode, and ions that are produced are  $[M+H]^+$  type ions. Acidified solvents are used to promote the protonation of the analyte. Nano-ESI, a miniaturized version of ESI, has the advantage of consuming very little sample, and its uses low flow rates to increase the ionization efficiency (Karas, Bahr et al. 2000).

### 1.2.2 The Mass Analyzer

The mass analyzers are used to separate ions by their mass-to-charge ratio. Because of their charges, the ions' position and trajectory in gas phase can be manipulated with magnetic and electric fields (Watson and Sparkman 2007). The mass spectrometer operates in a reduced pressure environment. This vacuum is mandatory to maintain the focusing capability of the analyzer, as collision with neutral molecules will lead to ion diffusion.

A diverse and versatile range of analyzers exists on the market, and the important characteristics are the resolving power and the mass range. Resolving power ( $R$ ) is the capacity for an instrument to separate two peaks and is defined by the following equation.  $M$  is the  $m/z$  of the peak and  $\Delta m$  is the difference in  $m/z$  between two adjacent peaks:

$$R = \frac{M}{\Delta m}$$

Equation 1: Resolving Power

High resolution instruments have the advantage of performing accurate mass measurement, and therefore enable the determination of potential elemental formulae within specific mass tolerances. High mass accuracy is achieved when the instrument is capable of well separating neighboring peaks (Gross 2004). Note that an increase in resolution is usually made at the cost of the sensitivity. Mass accuracy is typically expressed in ppm (parts per million) which corresponds to the difference between the observed and calculated mass divided by the molecular mass of the analyte. Proteomic analysis takes advantage of high resolution instruments by reducing the false positive rate in peptide assignments (Mann and Kelleher 2008). For instance, an accuracy of +/- 1ppm reduces by 99 % all possible peptide assignments compared to using nominal mass measurements (Zubarev, Hakansson et al. 1996).

### 1.2.2.1 The Orbitrap mass analyzer

The mass analyzer used in this study is the LTQ-Orbitrap (Figure 1.6), a hybrid instrument comprising a linear ion trap combined with a high resolution orbital trap (Scigelova and Makarov 2006). This hybrid instrument contains two analyzers that can detect ions: the LTQ most commonly is used to generate tandem mass spectra (MS/MS) while the orbitrap detects all ions of the survey MS scan at high resolution. The linear ion trap or linear trap quadrupole (LTQ) are capable of generating MS and MS/MS at high sensitivity, but with low mass accuracy. In the Orbitrap mass analyzer, the ion oscillates around an electrode, resulting in a frequency which is a function of its  $m/z$ . The Orbitrap's main advantage is high resolution ( $\sim 100,000$  at  $m/z$  400) enabling high mass accuracy measurements (1-2ppm) (Scigelova and Makarov 2006) while the LTQ offers high speed and sensitivity. By using high resolution MS, it is possible to detect the isotopic pattern of a given ion. The isotopic pattern provides information about the charge state ( $z$ ) of the ion.

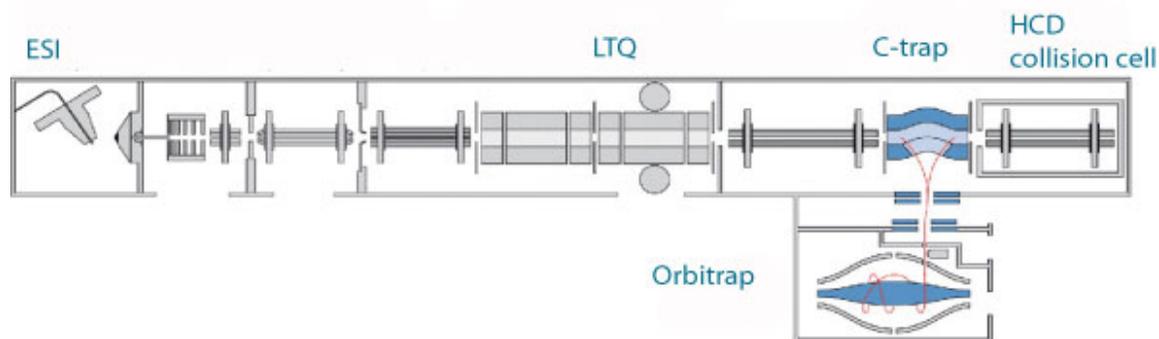


Figure 1.6: Peptides are ionized in the ESI and fragmented in either the LTQ or in the HCD collision cell. The Orbitrap is used to detect parent  $m/z$  (adapted from(Olsen, Schwartz et al. 2009))

The orbitrap is an ion trap, where ions are trapped around an electrode. It consists of 2 electrodes (inner and outer) between which an electric potential is imposed. Because of their shape, the electric field is non-homogenous between the electrodes; leading the

ions to oscillate along the inner electrode (Watson and Sparkman 2007). These oscillations are  $m/z$  dependent and this feature of the orbitrap makes it a mass analyzer.

This concept had existed for a long time, but the main challenge was to introduce ions in the orbitrap. The C trap was designed to focus ions before injecting them into small pockets and this process is coordinated with the increase in the electric field in the orbitrap (Watson and Sparkman 2007).

### **1.2.2.2 Linear ion trap mass analyzer**

The linear ion trap (LIT) is composed of a linear quadrupole and its function is to trap ions, select those of interest and fragment them (Watson and Sparkman 2007). High potential is applied at the front and the end of the quadrupole thus enabling the trapping. A potential well is created and the ions are stored in defined boundaries. To achieve a scan, the ions are ejected sequentially by applying an RF voltage and once ejected, the ions hit the detector. The ejection of the ion occurs when the RF voltage matches the frequency of the ion of interest which is  $m/z$  dependent (Gross 2004). The selection of a particular ion is achieved when all the ions of lower and higher  $m/z$  are ejected. When the ion of interest is isolated, collision induced dissociation can occur in the trap using helium gas, producing fragments that can be detected (Gross 2004).

### **1.2.3 Tandem MS**

MS/MS also referred to as tandem MS enables the acquisition of product ion spectra from precursor ions selected by the first mass analyzer. The second mass analyzer is thus scanned to transmit in turn fragment ions generated in a collision cell located between the two mass analyzers. The MS/MS spectrum reveals further structural information about the molecule and is the basis for peptide sequencing. Different algorithms currently exist to search the experimental MS/MS spectra and correlate observed fragment ions against those predicted from peptide candidates derived from a protein

database. Peptide fragmentation can be achieved using various techniques; some of which will be reviewed in this chapter.

Typically proteomics experiments involve the use of an LC-MS, and therefore ions are only observed at a given elution time in a short time window (30-60 s). Ideally all ions that are detected in a given MS scan should be fragmented and sequenced. Despite the short period for the acquisition of MS/MS spectra (~300 ns for CID), only a small fraction of all detected peptide ions can be sequenced in a given LC-MS/MS analysis. This method of collection is called data-dependent acquisition, where only a pre-selected number of the most intense ions are being selected for MS/MS (Mann, Hendrickson et al. 2001).

### **1.2.3.1 CID**

CID (collision induced dissociation) is by far the most popular fragmentation technique in proteomics. In CID, ions are first accelerated by an electrical potential, and then made to collide with neutral gas molecules such as nitrogen or helium. The collision between the incoming precursor ions and the target gas converts kinetic into potential energy that is distributed into the different oscillators of the ions. Bonds that have the lowest energy requirements (typically the amide bond of peptides under low energy CID) will be dissociated first. This technique is optimal when the precursor is doubly or triply charged, leading to singly charged fragments. The preferred fragmentation for peptides occurs at the amide bond, yielding b and y type fragments (refer to Figure 1.7). By looking at the mass difference of neighboring peaks, each of which corresponds to the mass of an amino acid, it is possible to deduce the peptide sequence. In the LTQ-Orbitrap instrument the CID fragmentation occurs in the LTQ.

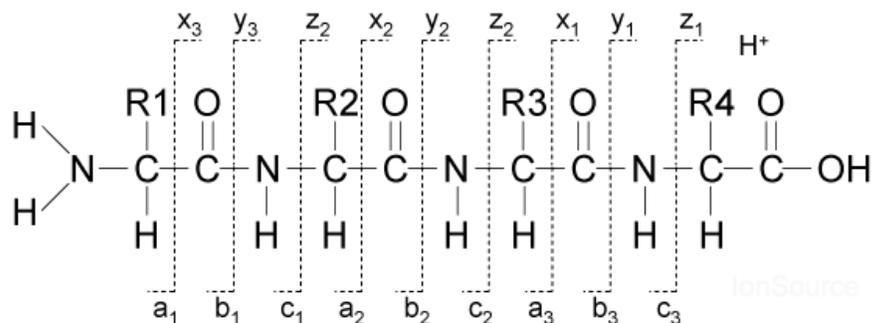


Figure 1.7: Nomenclature of fragment ions observed for the dissociation of peptide ions . Figure from: [http://www.ionsource.com/tutorial/DeNovo/full\\_anno.htm](http://www.ionsource.com/tutorial/DeNovo/full_anno.htm)

### 1.2.3.2ETD

ETD (electron transfer dissociation) is a dissociation technique that is based on adding a low energy electron to a multiply charged species (Watson and Sparkman 2007). The fragmentation is very specific to the N-C $\alpha$  bond (Boersema, Mohammed et al. 2009) first converting the peptide to a radical and forming c and z type fragment ions (see Figure 1.7). As electron source, ETD employs a radical anion (Syka, Coon et al. 2004) such as fluoranthene radical and the ETD-reaction is held in the LTQ. The fluoranthene has a low electron affinity; therefore it readily passes its electron to a peptide (Boersema, Mohammed et al. 2009). The main advantage of ETD is to fragment large peptides and those with labile PTMs (Syka, Coon et al. 2004). Contrary to CID, ETD gives best results with multiply charged species with  $z > 2$ .

### 1.2.3.3HCD

The fragmentation and detection in the ion trap is efficient and sensitive, however it lacks the mass resolution for the fragment ions and low  $m/z$  ions are not trapped (Olsen, Macek et al. 2007). In high energy collisional dissociation (HCD), the fragmentation occurs in an octopole collision cell located at the far end of the C-trap. The fragmentation mechanism and pattern is very similar to the LTQ CID fragmentation, leading to y and b ions. One advantage in HCD is the capability of detecting low  $m/z$

fragments. Since the detection occurs in the Orbitrap, high mass resolution is obtained on the product-ion spectra. Consequently, high mass resolution on fragment ions leads to higher confidence on the peptide identity.

#### **1.2.4 Detectors**

The last component of the mass spectrometer is the detector. Once the ions are separated, it is the detector that measures their respective signal. In the Orbitrap two detectors are present: a dynode detector after the LTQ and the image current measured from the motion of ions cycling in the Orbitrap analyzer. The dynode detector amplifies the signal by increasing the difference in potential in multiple steps. Once the ion hits the first electrode, multiple electrons are emitted which then hit the next electrode and so on (Gross 2004). In the Orbitrap the ions oscillate around a central electrode and surrounding plates record the frequency of the current followed by a Fourier transform that converts the frequency into  $m/z$  data (Makarov 2000).

The dynamic range is an important factor when considering the choice of a detector. The dynamic range is the ratio of the most intense peak over the least intense peak in the same spectrum (Gross 2004). In proteomics analysis the two main challenges are the sample complexity and the very high dynamic range of protein abundance in the cell. For comprehensive proteomic analysis, a high dynamic range is required; however, the present day dynamic range of  $10^3$ - $10^4$  is insufficient to cover the entire proteome (Mann and Kelleher 2008).

#### **1.2.5 Database Searching**

Following a single LC-MS run of 70 minutes, the instrument acquires up to 20 000 spectra. In a typical 2 condition proteomics experiment performed in this laboratory, about 800 000 spectra are generated. This large amount of data cannot be analyzed manually and computational methods have been developed to process the raw data.

The mass of the peptide is obtained from the MS while the tandem MS contains masses of fragments relevant to the sequence. It is possible to determine the peptide's sequence by calculating the mass difference between fragments and this procedure is called *de novo* sequencing. However, often MS/MS spectrum will only contain partial information and the entire sequence cannot be determined. In the 1990's, following genome sequencing, the peptide sequencing became a database-search problem. In nature, only a small number of combinations of amino acid sequence exists compared to all the possibilities when dealing with *de novo* sequencing. Nowadays, the most popular approach is through database searching, where MS and MS/MS scans are submitted to search engines such as MASCOT (Perkins, Pappin et al. 1999) or SEQUEST (Eng, McCormack et al. 1994). Thanks to the genome sequencing project, it is possible to virtually digest all the proteins present in an organism and create a peptide database. First, a database of peptides is generated for an entire genome, using the sequence of the proteins and the enzymatic cleavage rules (for instance, trypsin cuts at the N-terminus of lysine and arginine). The theoretical MS/MS spectrum for a specific fragmentation method is obtained for each of those peptides with different PTMs. The precursor mass (from the MS spectrum) and the fragment mass (from the MS/MS spectrum) of each theoretical peptide are compared to the experimental result. When submitting raw data to search engines, the first step is to use the peptide mass from the MS spectrum to obtain a list of all possible peptides with all allowed post-translational modifications respecting the enzyme cleavage rules that correspond to this mass within the allowed mass deviation. The second step is to generate mock MS/MS spectrum for those peptides, and a matching score is calculated for each every possibility. When a peptide is modified, the mass of the modified residue is considered when interpreting MS/MS spectrum. The comparison is done based on the allowed mass deviation for the precursor and the fragments' mass. For instance when acquiring data in the LTQ-Orbitrap the allowed mass deviation is around 15ppm for the precursor mass and 0.5Da for the fragments. Database-searching approaches can only be used for organisms' whose genome is sequenced and for which all the theoretical peptides are known (Steen and Mann 2004).

Each candidate peptide that matches to the experimental spectrum is assigned a score and a rank (Figure 1.8). The score is based on the quality of the match, for instance the number of fragments that were matched. Therefore, the score is the main parameter that is used to discriminate between right and wrong assignments. MASCOT, a widely used search-engine, uses the MOWSE algorithm to evaluate the match between the peptide and the spectrum (Pappin, Hojrup et al. 1993). MASCOT uses a probability-based scoring where the probability that a given match occurs randomly is calculated and its negative logarithm used as the score. The higher the score, the lower the probability that a given match is a random event. It is worth noting that the search engines assume modifications are unfragmentable during the CID or ETD process, which is true for small chemical modifications such as phosphorylation and acetylation; however this is not the case for the SUMO-modified peptide.

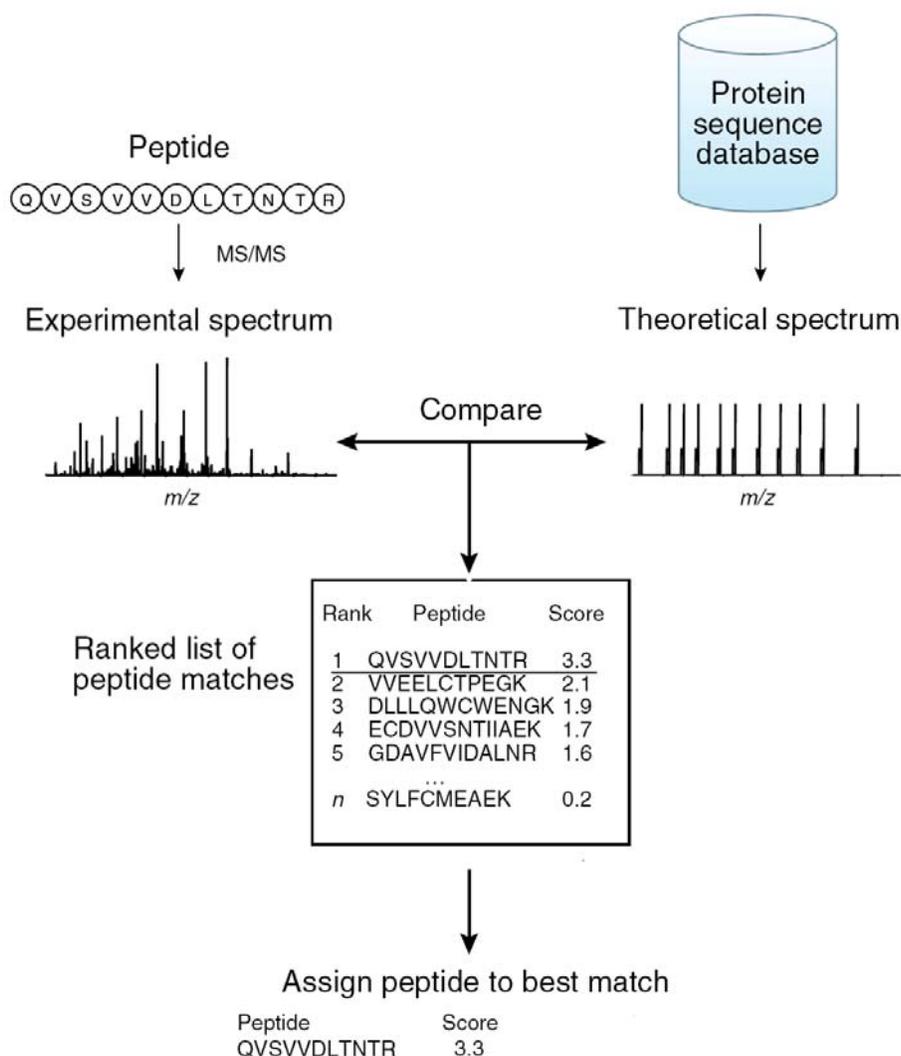


Figure 1.8: The MS/MS spectrum is compared against a theoretical MS/MS spectrum generated *in silico*. A score is given to each peptide based on the similarity of the two spectra. Adapted from (Nesvizhskii, Vitek et al. 2007).

The matching between the theoretical and the experimental peptide's MS/MS is not an ideal process and some errors occur (Elias and Gygi 2007). The low intensity of the precursor peptide, the poor quality of fragmentation, the fragmentation of two peptides in a single MS/MS all can lead to mistakes. In order to calculate the false discovery rate (FDR), the data is searched against a decoy database (Elias and Gygi 2007). The decoy database is created by reversing the protein sequences used in the database, and the MS/MS spectra are searched against a composite database: the forward and the reverse

database. At a given score cut-off, the FDR is calculated with the number of reverse-database matched peptides and therefore gives an estimate of the probability that a given match is a false positive.

Equation 2: Calculating the false discovery rate

$$FDR = 100 \times 2 \times \frac{\#reverse - database\ matched\ peptides}{Total\# peptides}$$

## 1.2.6 Quantitative Proteomics

Until recently, MS-based proteomics was mostly a qualitative technique that resulted in a list of proteins found under a given condition. In order to gain better insight into the biological relevance of the proteins present, the relative abundance of the proteins and of their respective modifications is necessary (Schulze and Usadel 2010). MS is not inherently quantitative: the ion's intensity not only depends on the ion's abundance, but also on the chemical properties of the peptide (charge, length, amino acid composition, etc.) and its environment (salts present during ionization). Absolute quantification cannot be performed on a given peptide simply based on its intensity. As a result, most of the large-scale MS quantification approaches always involve a comparison between two or more samples. Comparisons can only be made based on the same specie (the same peptide with the same  $m/z$ ) since different peptides have different ionization efficiencies.

A typical experimental design compares two conditions: a stressed *versus* a control status. Different quantification strategies have been developed in recent years and are divided as label-free and stable-isotope-labeling approaches (Schulze and Usadel 2010). In this laboratory, a label-free approach has been developed based on the correlation of peptide coordinates ( $m/z$ , time charge, intensity) across replicates and conditions. ProteoProfile is an in-house label-free bioinformatics software that quantifies peptides and proteins across different samples.

ProteoProfile creates peptide maps from LC-MS raw data and then clusters maps from different sample sets. First it creates contour maps that include retention time, charge, intensity and  $m/z$  for all detection peptide ions (see Figure 1.9).

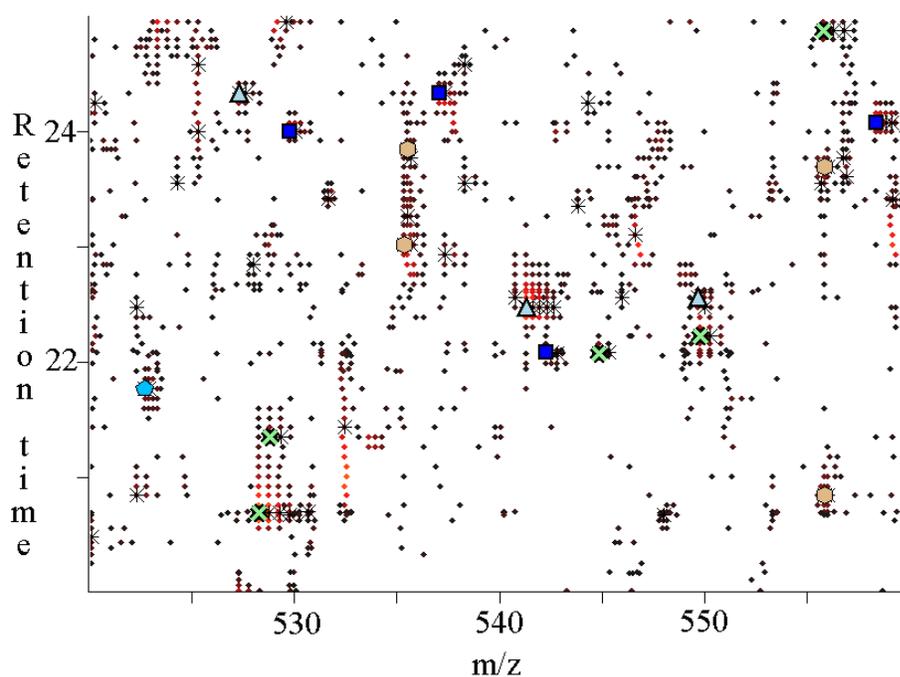


Figure 1.9: Contour map created by Proteoprofile. Symbols circle, cross and triangles are represent singly, doubly and triply charged species respectively, while the range of intensities is represented with color from dark red to yellow being the most intense.

Once maps are created for each sample, ProteoProfile then clusters the different samples together by aligning maps using linear dynamic correlation. Since the overlap in MASCOT identification between two LC-MS/MS of the same sample is around 60% due to the random process of selection when using data-dependent acquisition, the ions are not aligned based on their MASCOT identification, but rather accordingly to their retention time,  $m/z$  as well as the surrounding environment. Mass accuracy becomes even more important when dealing with label free quantification, since the clustering

relies on mass accuracy. Therefore, high-mass resolution instruments made label-free quantification even more appealing in recent years (Schulze and Usadel 2010).

### 1.3 SUMO MS based proteomics

In this section, different enrichment techniques of SUMO-conjugated proteins and SUMO-modified peptides bearing the modification sites will be described, followed by spectral interpretation strategies.

#### 1.3.1 Enrichment of SUMOylated proteins

In order to identify Ub/Ubl modified proteins and their sites, the need for an enrichment strategy became apparent due to the low abundance and high turnover of those species. A strategy was successfully employed to identify ubiquitin-conjugated proteins and their modification sites in yeast: around 1000 potential conjugates and 100 ubiquitylation sites were identified (Peng, Schwartz et al. 2003).

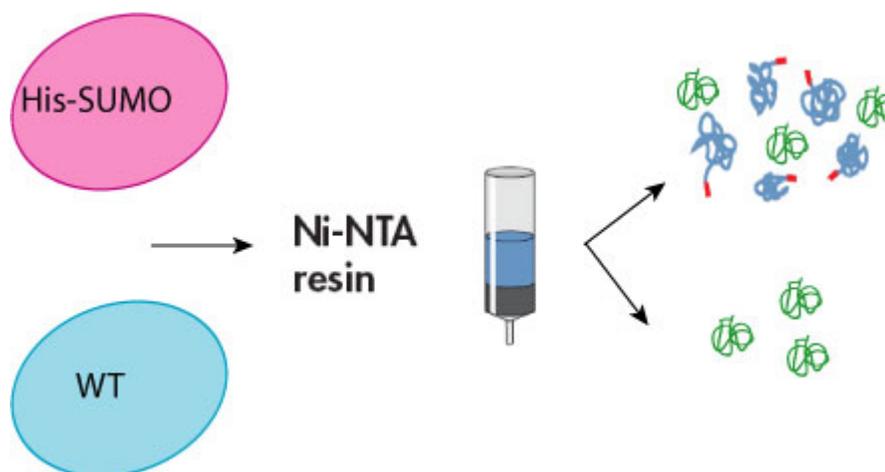


Figure 1.10: Purification strategy of His-SUMO conjugated proteins using the Ni-NTA pull down. Contaminants are identified using the WT control and the SUMO-conjugated proteins are uniquely present in the His-SUMO sample.

The approach described in Figure 1.10 and developed by Gygy laboratory is based on the expression of His<sub>6</sub>-ubiquitin in yeast and its subsequent purification. His<sub>6</sub>-ubiquitin-conjugated proteins are isolated using Ni-NTA (nickel nitriloacetic acid) resin in denaturing conditions. In parallel, a control sample (wild type) expressing no His<sub>6</sub>-Ub is purified in the same manner and the isolated proteins are the non-specific contaminants that bind the resin. The success of this proteomic study inspired multiple groups to study SUMO-conjugates in a similar manner in yeast and in mammals (Matic, Schimmel et al. 2010), (Blomster, Imanishi et al. 2010), (Wohlschlegel, Johnson et al. 2004), (Wohlschlegel, Johnson et al. 2006). The denaturing conditions are extremely important, since they first assure the inactivity of SENPs, and secondly assure the exclusive purification of SUMO-conjugates proteins and not their associated proteins. The proteins that are uniquely identified in the His<sub>6</sub>-SUMO expressing strain are assumed to be *bona fide* SUMO substrates. Up to now hundreds of putative SUMO conjugated proteins have been identified.

### 1.3.2 Enrichment strategies of SUMO-modified peptides

Mass spectrometric analysis of SUMOylation sites is technically very challenging and only a few sites have been identified until recently (Wilson and Heaton 2008). The majority of identified SUMO targets are only putative targets and their modification sites are still unknown. The identification of the conjugation sites can be important in understanding the regulation of SUMOylation.

The digestion of purified SUMO conjugates leads to a mixture of peptides, most of which do not contain the modification site. This mixture is too complex for efficient identification of the modification site. Since SUMO-modified proteins are very low in abundance, this is even more problematic.

Secondly, following trypsin digestion, the modified peptide is a branched peptide that contains a large C-terminal SUMO sequence. The absence of arginine and lysine residues makes the chromatographic separation and the MS analysis challenging.

Different strategies have been designed to alleviate the MS interpretation and SUMO peptide enrichment.

In order to purify SUMO1-modified peptides, Bloomster et al engineered the human SUMO-1 protein so that following LysC digestion, the resulting peptide contains two cysteines followed by an arginine on the SUMO1 moiety (Blomster, Imanishi et al. 2010). The SUMO1-modified peptides are then enriched using thiopropyl sepharose beads that bind through disulfide linkage with the SUMO1-modified peptides. The peptides of interest are then released from the beads using trypsin digestion that leaves a di-glycine tag on the lysine. This method allowed the identification of 14 SUMOylation sites in HeLa cells (Blomster, Imanishi et al. 2010). However, this method presents several drawbacks, such as the difficulty of distinguishing SUMO1-modified from Ubiquitin-modified peptides. The second drawback is the possible enrichment of non-SUMO1-modified peptides that contain cysteine residues.

A second purification strategy was developed using an engineered SUMO2/3. A LysC resistant His-SUMO protein was engineered where all lysines residues were mutated to arginines (Matic, Schimmel et al. 2010). The first step involves a LysC digestion followed by the Ni-NTA enrichment of SUMO protein linked to Lys-C digested target peptide. Following the Ni-NTA enrichment, the mixture is digested by trypsin. Using this method, 103 SUMO-2/3 acceptor sites were recently identified in mammals (Matic, Schimmel et al. 2010). Since all lysine residues are substituted by arginine, this mutant SUMO2 cannot form polySUMOylation and thus changes the function of the corresponding protein.

All modified peptide based enrichment strategies present a highly modified form of the SUMO protein and therefore can introduce some artifacts.

### 1.3.3 Stress-induced SUMOylation

Numerous studies observed an accumulation of SUMO-conjugates upon cellular stress (Saitoh and Hinchev 2000; Kurepa, Walker et al. 2003; Bohren, Nadkarni et al. 2004; Manza, Codreanu et al. 2004; Golebiowski, Matic et al. 2009). Various external stress such as heat shock, oxidative and genotoxic chemicals have been tested on the SUMO system in many cell types, and SUMO4 expression has only been reported in extreme stress conditions (Wei, Yang et al. 2008). This effect is mostly observed with SUMO2/3 conjugates, following stress, the free SUMO2/3 is highly conjugated (Saitoh and Hinchev 2000), however SUMO1 does not seem to be affected in the same manner.

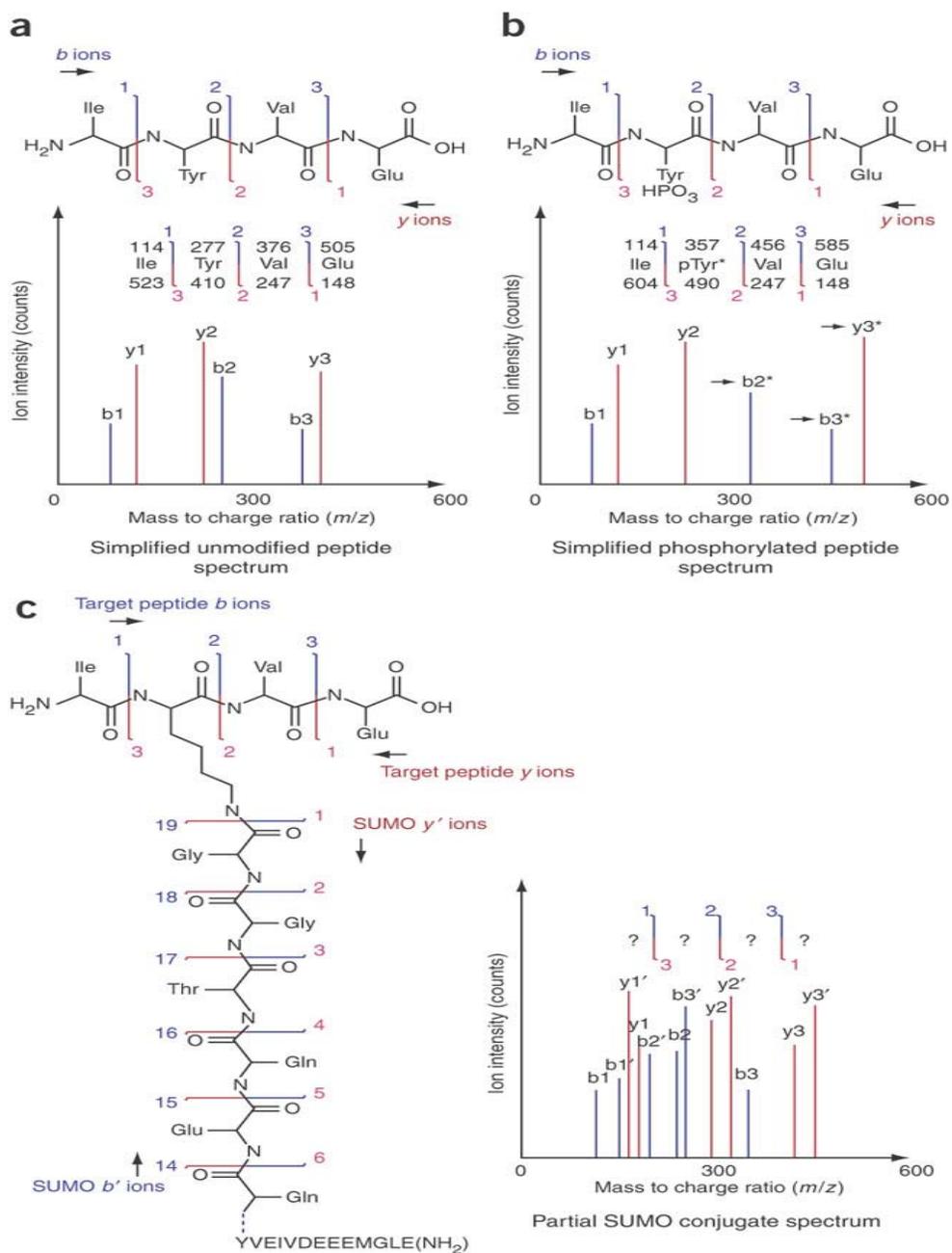
When exposed to arsenic trioxide, PML and PML-RAR $\alpha$  are SUMOylated and subsequently degraded by the ubiquitin-proteasome system (Lallemand-Breitenbach, Jeanne et al. 2008). PML-RAR $\alpha$  is a fusion protein expressed in APL (acute promyelocytic leukaemia) and As<sub>2</sub>O<sub>3</sub> is an agent used to treat APL (Zhu, Chen et al. 2002).

### 1.3.4 Spectral interpretation strategies

So far, the common strategy in identifying SUMOylation sites is through site-directed mutagenesis of consensus site lysines. However, it has been shown that non-consensus lysines can be SUMOylated as for E2-25K (Pichler, Knipscheer et al. 2005) and the opposite is also true: not all consensus site lysine are SUMOylated. In order to gain a better insight on the biological role of SUMOylation, the need for a new and unbiased strategy for SUMO site identification has become important in the last decade. Mass spectrometry is the method of choice in identifying the modification sites for numerous PTMs.

Most PTMs, such as ubiquitylation, methylation or acetylation, do not fragment during the CID or the ETD process in the MS. As an example, phosphorylation sites can

be readily identified using standard sequencing softwares by including phosphorylation as a variable modification and simply altering the fragment's mass by the mass of the modification (Figure 1.11b). Following trypsin digestion, the large SUMO tags produce multiple fragments that lead to complex fragmentation patterns, and although this adds confidence to the identification of the SUMO moiety, the MS/MS spectra is highly complex (Figure 1.11c) and renders spectral interpretation by MASCOT more prone to errors. Currently, two softwares were developed to specifically address this issue: SUMmOn (Pedrioli, Raught et al. 2006) and ChopNSpice (Hsiao, Meulmeester et al. 2009). SUMmOn algorithm was developed to interpret complex fragmentation patterns of PTMs such as SUMO (Pedrioli, Raught et al. 2006). The CID spectra are scanned for the specific fragments that originate from the modifier. SUMmOn calculates two scores one for the target peptide and one for the modification, and although this system seemed to be successful for *in vitro* by identifying RanGap SUMOylation site, no conclusive result has been observed *in vivo* on a large scale data sets.



The branched SUMO-modified peptides fragment in a manner similar to that of the linear peptide with a lysine missed cleavage and the SUMO moiety on the N-terminus (Hsiao, Meulmeester et al. 2009). Based on this observation, ChopNSpice algorithm was developed. A database is constructed where every possible lysine is modified by SUMO and the resulting peptide is linearized. The MS and MS/MS spectra are searched against the ChopNSpice generated database and using this strategy 18 SUMOylation sites were identified *in vivo* for SUMO-1 in HeLa cells (Hsiao, Meulmeester et al. 2009).

Although new bioinformatics strategies are more suitable for studying SUMOylation by MS, further technological advances would be required to provide more comprehensive SUMOylation analyses.

## 1.4 Project's goal

SUMO are a group of ~10kDa proteins that form an isopeptide linkage with the  $\epsilon$  amino group of the target lysine. SUMOylation is a post-translational modification required for cell viability (Johnson, Schwienhorst et al. 1997; Hayashi, Seki et al. 2002) and shares great similarity with ubiquitin (Bayer, Arndt et al. 1998). At present, around a hundred SUMO targets have been identified and the molecular consequences of protein SUMOylation *in vivo* such as reciprocity between ubiquitylation and SUMOylation, degradation and transcription regulation are unpredictable (reviewed in (Geiss-Friedlander and Melchior 2007)). This modification affects protein function in a diverse, complex and sometimes opposite way.

Large scale proteomics studies have been successful in identifying a high number of potential SUMO targets (Li, Evdokimov et al. 2004; Vertegaal, Ogg et al. 2004; Wohlschlegel, Johnson et al. 2004; Denison, Rudner et al. 2005; Gocke, Yu et al. 2005; Ganesan, Kho et al. 2007; Flick and Kaiser 2009). However, SUMOylation site identification is still a challenge, due to the very low abundance of SUMO targets *in vivo* as well as its rapid turnover upon cell lysis. Until now, SUMOylation site identification has mainly relied on site-directed mutagenesis, a time-consuming and burdensome task. To get a better insight into the molecular function of SUMOylation and a confidence in the target's identity, the number and the location of the exact site is essential.

In this context, the goal of this project is to develop a large scale MS based method to identify SUMOylation targets and their respective sites. Because of the very low abundance of SUMOylated proteins in the cell, and its highly dynamic nature (Geiss-Friedlander and Melchior 2007), the first part of this thesis will focus on target enrichment.

Following trypsin digestion of SUMOylated proteins, the peptides containing the modification site are branched. The analysis of branched peptide by MS is very limited,

partly due to the complexity of the MS-MS spectrum and the absence of efficient sequencing software (Pedrioli, Raught et al. 2006; Hsiao, Meulmeester et al. 2009). Hence, the second goal of this project is based on the interpretation of SUMO peptide MS/MS spectra.

## 1.5 References

- Ayaydin, F. and M. Dasso (2004). "*Distinct in vivo dynamics of vertebrate SUMO paralogues.*" Mol Biol Cell **15**(12): 5208-18.
- Bayer, P., A. Arndt, et al. (1998). "*Structure determination of the small ubiquitin-related modifier SUMO-1.*" Journal of Molecular Biology **280**(2): 275-286.
- Bernier-Villamor, V., D. A. Sampson, et al. (2002). "*Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1.*" Cell **108**(3): 345-56.
- Blomster, H. A., S. Y. Imanishi, et al. (2010). "*In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification.*" J Biol Chem **285**(25): 19324-9.
- Boersema, P. J., S. Mohammed, et al. (2009). "*Phosphopeptide fragmentation and analysis by mass spectrometry.*" Journal of Mass Spectrometry **44**(6): 861-878.
- Braschi, E., R. Zunino, et al. (2009). "*MAPL is a new mitochondrial SUMO E3 ligase that regulates mitochondrial fission.*" EMBO Rep **10**(7): 748-754.
- Cooper, H. J., M. H. Tatham, et al. (2005). "*Fourier transform ion cyclotron resonance mass spectrometry for the analysis of small ubiquitin-like modifier (SUMO) modification: identification of lysines in RanBP2 and SUMO targeted for modification during the E3 autoSUMOylation reaction.*" Anal Chem **77**(19): 6310-9.
- Denison, C., A. D. Rudner, et al. (2005). "*A proteomic strategy for gaining insights into protein sumoylation in yeast.*" Mol Cell Proteomics **4**(3): 246-54.

- Elias, J. E. and S. P. Gygi (2007). "*Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.*" Nat Methods **4**(3): 207-14.
- Eng, J. K., A. L. McCormack, et al. (1994). "*An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.*" Journal of the American Society for Mass Spectrometry **5**(11): 976-989.
- Fenn, J. B., M. Mann, et al. (1989). "*Electrospray ionization for mass spectrometry of large biomolecules.*" Science **246**(4926): 64-71.
- Flick, K. and P. Kaiser (2009). "*Proteomic Revelation: SUMO Changes Partners When the Heat Is On.*" Sci. Signal. **2**(81): pe45-.
- Ganesan, A. K., Y. Kho, et al. (2007). "*Broad spectrum identification of SUMO substrates in melanoma cells.*" Proteomics **7**(13): 2216-21.
- Geiss-Friedlander, R. and F. Melchior (2007). "*Concepts in sumoylation: a decade on.*" Nat Rev Mol Cell Biol **8**(12): 947-56.
- Gocke, C. B., H. Yu, et al. (2005). "*Systematic Identification and Analysis of Mammalian Small Ubiquitin-like Modifier Substrates.*" J. Biol. Chem. **280**(6): 5004-5012.
- Golebiowski, F., I. Matic, et al. (2009). "*System-wide changes to SUMO modifications in response to heat shock.*" Sci Signal **2**(72): ra24.
- Gross, H. J. (2004). *Mass Spectrometry*. Berlin Heidelberg, Springer-Verlag.

- Guo, D., M. Li, et al. (2004). "*A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes.*" Nat Genet **36**(8): 837-41.
- Hayashi, T., M. Seki, et al. (2002). "*Ubc9 Is Essential for Viability of Higher Eukaryotic Cells.*" Experimental Cell Research **280**(2): 212-221.
- Hietakangas, V., J. Anckar, et al. (2006). "*PDSM, a motif for phosphorylation-dependent SUMO modification.*" Proc Natl Acad Sci U S A **103**(1): 45-50.
- Hsiao, H. H., E. Meulmeester, et al. (2009). "*ChopNSpice, a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides.*" Mol Cell Proteomics **8**(12): 2664-75.
- Kamitani, T., K. Kito, et al. (1998). "*Identification of three major sentrinization sites in PML.*" J Biol Chem **273**(41): 26675-82.
- Karas, M., U. Bahr, et al. (2000). "*Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine.*" Fresenius J Anal Chem **366**(6-7): 669-76.
- Karas, M. and F. Hillenkamp (1988). "*Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.*" Anal Chem **60**(20): 2299-301.
- Kerscher, O., R. Felberbaum, et al. (2006). "*Modification of proteins by ubiquitin and ubiquitin-like proteins.*" Annu Rev Cell Dev Biol **22**: 159-80.
- Kurepa, J., J. M. Walker, et al. (2003). "*The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress.*" J Biol Chem **278**(9): 6862-72.

- Lallemand-Breitenbach, V. and H. de The (2010). "*PML nuclear bodies.*" Cold Spring Harb Perspect Biol **2**(5): a000661.
- Lallemand-Breitenbach, V., M. Jeanne, et al. (2008). "*Arsenic degrades PML or PML-RAR $\alpha$  through a SUMO-triggered RNF4/ubiquitin-mediated pathway.*" Nat Cell Biol **10**(5): 547-55.
- Mahajan, R., C. Delphin, et al. (1997). "*A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2.*" Cell **88**(1): 97-107.
- Makarov, A. (2000). "*Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.*" Anal Chem **72**(6): 1156-62.
- Mann, M., R. C. Hendrickson, et al. (2001). "*Analysis of proteins and proteomes by mass spectrometry.*" Annual Review of Biochemistry **70**(1): 437-473.
- Mann, M. and N. L. Kelleher (2008). "*Precision proteomics: The case for high resolution and high mass accuracy.*" Proceedings of the National Academy of Sciences **105**(47): 18132-18138.
- Manza, L. L., S. G. Codreanu, et al. (2004). "*Global shifts in protein sumoylation in response to electrophile and oxidative stress.*" Chem Res Toxicol **17**(12): 1706-15.
- Marcin Drag, G. S. S. (2008). "*DeSUMOylating enzymes - SENPs.*" IUBMB Life **60**(11): 734-742.

- Martin, S., K. A. Wilkinson, et al. (2007). "*Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction.*" Nat Rev Neurosci **8**(12): 948-59.
- Matic, I., J. Schimmel, et al. (2010). "*Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif.*" Mol Cell **39**(4): 641-652.
- Matic, I., M. van Hagen, et al. (2008). "*In Vivo Identification of Human Small Ubiquitin-like Modifier Polymerization Sites by High Accuracy Mass Spectrometry and an in Vitro to in Vivo Strategy.*" Mol Cell Proteomics **7**(1): 132-144.
- Mukhopadhyay, D. and M. Dasso (2007). "*Modification in reverse: the SUMO proteases.*" Trends Biochem Sci **32**(6): 286-95.
- Nesvizhskii, A. I., O. Vitek, et al. (2007). "*Analysis and validation of proteomic data generated by tandem mass spectrometry.*" Nat Meth **4**(10): 787-797.
- Owerbach, D., E. M. McKay, et al. (2005). "*A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation.*" Biochem Biophys Res Commun **337**(2): 517-20.
- Pappin, D. J. C., P. Hojrup, et al. (1993). "*Rapid identification of proteins by peptide-mass fingerprinting.*" Current biology : CB **3**(6): 327-332.
- Pedrioli, P. G., B. Raught, et al. (2006). "*Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software.*" Nat Methods **3**(7): 533-9.

- Peng, J., D. Schwartz, et al. (2003). "*A proteomics approach to understanding protein ubiquitination.*" Nat Biotech **21**(8): 921-926.
- Perkins, D. N., D. J. Pappin, et al. (1999). "*Probability-based protein identification by searching sequence databases using mass spectrometry data.*" Electrophoresis **20**(18): 3551-67.
- Petsko, G. A. and D. Ringe *Protein Structure and Function*, New Science Press.
- Pichler, A., P. Knipscheer, et al. (2005). "*SUMO modification of the ubiquitin-conjugating enzyme E2-25K.*" Nat Struct Mol Biol **12**(3): 264-9.
- Pickart, C. M. and D. Fushman (2004). "*Polyubiquitin chains: polymeric protein signals.*" Current Opinion in Chemical Biology **8**(6): 610-616.
- Reineke, E. L., Y. Liu, et al. (2009). "*Promyelocytic leukemia protein controls cell migration in response to hydrogen peroxide and insulin-like growth factor-1.*" J Biol Chem **285**(13): 9485-92.
- Rodriguez, M. S., C. Dargemont, et al. (2001). "*SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting.*" Journal of Biological Chemistry **276**(16): 12654-12659.
- Saitoh, H. and J. Hinchev (2000). "*Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3.*" J Biol Chem **275**(9): 6252-8.
- Schulze, W. X. and B. r. Usadel (2010). "*Quantitation in Mass-Spectrometry-Based Proteomics.*" Annual Review of Plant Biology **61**(1): 491-516.

- Scigelova, M. and A. Makarov (2006). "*Orbitrap Mass Analyzer – Overview and Applications in Proteomics.*" Proteomics **6**(S2): 16-21.
- Semple, C. A. (2003). "*The comparative proteomics of ubiquitination in mouse.*" Genome Res **13**(6B): 1389-94.
- Steen, H. and M. Mann (2004). "*The abc's (and xyz's) of peptide sequencing.*" Nat Rev Mol Cell Biol **5**(9): 699-711.
- Syka, J. E. P., J. J. Coon, et al. (2004). "*Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.*" Proceedings of the National Academy of Sciences of the United States of America **101**(26): 9528-9533.
- Tatham, M. H., E. Jaffray, et al. (2001). "*Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9.*" J Biol Chem **276**(38): 35368-74.
- Ulrich, H. D. (2008). The SUMO System: An Overview. **497**: 3-16.
- Van Damme, E., K. Laukens, et al. (2010). "*A manually curated network of the PML nuclear body interactome reveals an important role for PML-NBs in SUMOylation dynamics.*" Int J Biol Sci **6**(1): 51-67.
- Vertegaal, A. C., S. C. Ogg, et al. (2004). "*A proteomic study of SUMO-2 target proteins.*" J Biol Chem **279**(32): 33791-8.
- Voet, D. and J. G. Voet (1995). *Biochemistry*, John Wiley & Sons, inc.

- Watson, T. J. and D. O. Sparkman (2007). *Introduction, in Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation*. Chichester, UK, John Wiley & Sons, Ltd.
- Weisshaar, S. R., K. Keusekotten, et al. (2008). "Arsenic trioxide stimulates SUMO-2/3 modification leading to RNF4-dependent proteolytic targeting of PML." FEBS Lett **582**(21-22): 3174-8.
- Wilson, V. G. and P. R. Heaton (2008). "Ubiquitin proteolytic system: focus on SUMO." Expert Rev Proteomics **5**(1): 121-35.
- Wohlschlegel, J. A., E. S. Johnson, et al. (2006). "Improved identification of SUMO attachment sites using C-terminal SUMO mutants and tailored protease digestion strategies." J Proteome Res **5**(4): 761-70.
- Xue, Y., F. Zhou, et al. (2006). "SUMOsp: a web server for sumoylation site prediction." Nucleic Acids Res **34**(Web Server issue): W254-7.
- Yang, S. H., A. Galanis, et al. (2006). "An extended consensus motif enhances the specificity of substrate modification by SUMO." Embo J **25**(21): 5083-93.
- Zhu, J., Z. Chen, et al. (2002). "How acute promyelocytic leukaemia revived arsenic." Nat Rev Cancer **2**(9): 705-13.
- Zubarev, R. A., P. Hakansson, et al. (1996). "Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements." Analytical Chemistry **68**(22): 4060-4063.

## **2. A Novel Proteomics Approach to Identify SUMOylated Proteins and their Modification Sites in Human Cells**

Frederic Galisson<sup>#</sup>, Louiza Mahrouche<sup>#</sup>, Mathieu Courcelles, Eric Bonneil, Sylvain Meloche, Mounira K. Chelbi-Alix, Pierre Thibault

(2010) *Molecular & Cellular Proteomics*, published.

<sup>#</sup>These authors contributed equally to this work

Louiza Mahrouche : Developed immunoprecipitation method, performed MS experiments and data analysis.

Frederic Galisson : created the mutant SUMOs and stable cell lines. Developed Ni-NTA purification and data analysis.

Mathieu Courcelle : developed informatics tools interpreting SUMO MS/MS spectra.

Eric Bonneil : Supported the MS experiments.

Sylvain Meloche : Original idea

Mounira K. Chelbi-Alix : Original idea and supervision

Pierre Thibault : Original idea and supervision

## 2.1 Abstract

The small ubiquitin-related modifier (SUMO) represents a small group of proteins that are reversibly attached to protein substrates to modify their functions. The large-scale identification of protein SUMOylation and their modification sites in mammalian cells represent a significant challenge due to the relatively small number of *in vivo* substrates and the dynamic nature of this modification. We report here a proteomics approach to selectively enrich and identify SUMO conjugates from human cells. HEK293 cells stably expressing the different SUMO paralogs containing a His<sub>6</sub> tag and a strategically located tryptic cleavage site to facilitate the recovery, identification and the isolation of SUMOylated peptides by affinity enrichment and mass spectrometry. The formation of tryptic peptides with short SUMO remnants offer significant advantages in large-scale SUMOylome experiments by generating paralog-specific fragment ions following CID and ETD activation and facilitating the identification of modified peptides using conventional database search engines such as Mascot. For SUMO-3 extracts, we identified 205 unique protein substrates together with 17 precise SUMOylation sites present in 12 SUMO protein conjugates using single-step NTA enrichment. Amongst SUMOylated substrates we identified promyelocytic leukemia (PML) with three new sites (K380, K400 and K497). The combination of a sequential NTA and immunoaffinity enrichment enabled the identification of 3 SUMO-1 modified peptides on low abundance SUMO-1 substrates (RanGAP1, SAFB2 and PML). Label-free quantitative proteomics analyses on untreated and arsenic trioxide-treated cells revealed that all identified SUMOylated sites of PML for both SUMO-1 and SUMO-3 were differentially SUMOylated upon stimulation.

## 2.2 Introduction

The small ubiquitin-like modifier (SUMO) proteins are structurally similar to ubiquitin although they share less than 20% sequence identity (Kerscher, Felberbaum et al. 2006). Like ubiquitylation, protein SUMOylation is regulated by a cascade of reactions involving SUMO-activating enzymes (SAE1/SAE2), conjugating enzymes (Ubc9) and one of several SUMO-E3 ligases (e.g. PIAS1, PIAS3, PIASx $\alpha$ , PIASx $\beta$ , PIASy, RanBP2 and Pc2) that covalently attach SUMO to specific protein substrates (Guo, Yang et al. 2007; Hsiao, Meulmeester et al. 2009). SUMO proteins are expressed as an immature proform that comprise an invariant Gly-Gly motif followed by a C terminal stretch of variable length (2–11 amino acids). Removal of this C terminal extension by sentrin-specific proteases (SENPs) to expose the di-glycine motif is necessary for the conjugation of SUMO to protein targets. These SUMO proteases are able to cleave both a peptide bond during the formation of mature SUMO, and an isopeptide bond to deconjugate modified protein substrates (Mukhopadhyay and Dasso 2007). This covalent modification arises from the formation of an isopeptide bond between the  $\epsilon$ -amino group of a lysine within the protein substrate and the C-terminus carboxy group of the SUMO glycine residue. SUMO conjugation frequently occurs at the lysine residue within the consensus motif  $\psi$ KxE (where  $\psi$  is an aliphatic residue and x any amino acid) that is recognized by Ubc9 (Bernier-Villamor, Sampson et al. 2002; Lin, Tatham et al. 2002). Recent studies have also identified a phosphorylation-dependent motif ( $\Psi$ KxE<sub>xxp</sub>SP) (Hietakangas, Anckar et al. 2006) and a negatively charged amino-acid-dependent motif (Yang, Galanis et al. 2006), that harbor negative charges next to the basic SUMO consensus site to enhance protein SUMOylation. However, several other SUMOylated proteins including PCNA, E2-25K, Daxx, and USP25 are modified at non-consensus sites (Hoegel, Pfander et al. 2002; Pichler, Knipscheer et al. 2005; Shen, Tatham et al. 2006). Whether these types of sites are rare exceptions or reflect the presence of other E2-conjugating enzymes is presently unknown.

In lower eukaryotes, a single SUMO gene is expressed (*Smt3* in *Saccharomyces cerevisiae*), whereas in vertebrates three paralogs designated as SUMO-1, SUMO-2 and SUMO-3 are ubiquitously expressed in all tissues. The human genome also encodes a fourth gene for SUMO-4 that seems to be uniquely expressed in the spleen, lymph nodes and kidney (Guo, Li et al. 2004). However, its role remains enigmatic, as its *in vivo* maturation into a conjugation-competent form still remains unclear (Owerbach, McKay et al. 2005). Interestingly, SUMO-2 and SUMO-3 share 97% sequence identity, and are expressed at much higher levels than SUMO-1, with which they only share about 50% identity (Kerscher, Felberbaum et al. 2006). Although SUMO paralogs use the same conjugation machinery and have partial overlapping subsets of target proteins, they respond differently to stress (Golebiowski, Matic et al. 2009) and can be distinguished by their ability to form self-modified polymers *in vivo* and *in vitro* (Tatham, Jaffray et al. 2001; Matic, van Hagen et al. 2008). SUMO-1 lacks a consensus modification site and does not form polySUMO-1 chains *in vivo*, although RanBP2 was reported to be hypermodified by SUMO-1 chains *in vitro* (Pichler, Gast et al. 2002). In contrast, SUMO-2 and SUMO-3 can form polymeric chains *in vivo* and *in vitro* through their consensus motif (Tatham, Jaffray et al. 2001), whereas SUMO-1 forms terminating chain on poly-SUMO-2 or poly-SUMO-3 conjugates (Matic, van Hagen et al. 2008).

Protein SUMOylation is an essential cellular process conserved from yeast to mammals and plays an important role in the regulation of intracellular trafficking, cell cycle, DNA repair and replication, cell signaling and stress responses (Hay 2005; Bossis and Melchior 2006; Hsiao, Meulmeester et al. 2009). Protein SUMOylation imparts significant structural and conformational changes on the substrate proteins by masking and/or by conferring additional scaffolding surfaces for protein interactions. At present, a few hundred protein substrates are known to be SUMOylated *in vivo*. These protein targets include regulators of gene expression (e.g. transcription factors, co-activators or repressors) as well as oncogenes and tumor suppressor genes, such as promyelocytic leukaemia (PML), Mdm2, c-Myb, c-Jun, and p53 whose misregulation leads to tumorigenesis and metastasis (Hoeller and Dikic 2009). There is growing evidences of cross-talk between protein SUMOylation and ubiquitylation processes (Denuc and

Marfany ; Schimmel, Larsen et al. 2008). Earlier reports indicated that SUMOylation can antagonize the ubiquitylation of NFκB (Desterro, Rodriguez et al. 1998) whereas recent data also suggest that SUMOylation can be a prerequisite for ubiquitylation and subsequent proteasome-dependent degradation. A case in point is the identification of RNF4, an E3 ubiquitin ligase that specifically recognizes and ubiquitinylates polySUMO-chains of PML (Lallemand-Breitenbach, Jeanne et al. 2008; Tatham, Geoffroy et al. 2008). Interestingly, PML SUMOylation can also be enhanced using arsenic trioxide (As<sub>2</sub>O<sub>3</sub>), a therapeutic agent used for the treatment of acute promyelocytic leukemia (APL) (Zhu, Koken et al. 1997; Muller, Matunis et al. 1998; Lallemand-Breitenbach, Jeanne et al. 2008).

The relatively low abundance of protein SUMOylation is a significant analytical challenge for the identification and quantitation of this modification *in vivo*. Recent reports have described the successful identification of SUMO protein candidates by enriching the small subset of SUMOylated proteins using cysteine-targeted purification (Blomster, Imanishi et al. 2010), tandem affinity tag with His<sub>6</sub>-SUMO proteins resistant to Lys C proteolysis (Matic, Schimmel et al. 2010), cells stably expressing His<sub>6</sub>-SUMO constructs (Tatham, Rodriguez et al. 2009) and quantifying their proportions using mass spectrometry (MS) and metabolic labelling in cell cultures (Vertegaal, Andersen et al. 2006). In spite of these significant advances, the identification of SUMOylation sites by MS remains challenging due to their variable stoichiometry and the presence of long SUMO C-termini polypeptides that complicates the interpretation of the corresponding product ion spectra. Upon tryptic digestion, SUMOylated peptides contain a relatively long SUMO remnant chain (up to 32 amino acid for human SUMO-2,3) appended on the modified Lys side chain. In contrast to other protein modifications, the SUMO remnant chain gives rise to multiple fragment ions that overlap with those of the target peptide. Accordingly, standard database search engines tailored to identify linear peptides are generally not capable of assigning the correct sequence of the branched peptides due to the complex distribution of overlapping fragment ions. This limitation was also described in an earlier report by Wohlschlegel et al. who indicated that yeast *Smt3* mutant with an arginine at the 3<sup>rd</sup> residue from the C-terminus yielded tryptic peptides

identical to those of ubiquitin remnant and facilitated their identification using the database engine SEQUEST (Wohlschlegel, Johnson et al. 2006). To overcome some of these limitations, different database searching strategies including an automated recognition pattern tool (SUMmOn) (Pedrioli, Raught et al. 2006) and combined search engines (ChopNSpice) (Hsiao, Meulmeester et al. 2009) were developed to identify potential acceptor sites from MS/MS spectra of these large precursor ions.

Here, we present a new approach to the large-scale identification of modified peptides and their conjugation sites using three separate HEK293 cell lines, each stably expressing a SUMO mutant protein that contain a His<sub>6</sub> tag and a strategically located tryptic cleavage site enabling convenient affinity enrichment and MS analyses of distinct SUMO paralogs. We introduced an Arg residue at the 6<sup>th</sup> position from the C-terminus to minimize structural changes with respect to the endogenous proteins. We confirmed the activity and functional properties of the His<sub>6</sub>-SUMO mutants using *in vitro* SUMOylation assays and immunofluorescence. We profiled the subcellular distribution of SUMOylated proteins in HEK293 cells and monitored the change in SUMOylation upon As<sub>2</sub>O<sub>3</sub> treatment. MS analyses of NTA-enriched nuclear cell extracts from mock and His<sub>6</sub>-SUMO-1/3 mutant HEK293 cells enabled the identification of unique SUMOylated protein substrates and their modification sites, including three novel sites on the protein PML. The separation of His<sub>6</sub>-containing proteins selectively enriched the mutated SUMOylated proteins from the endogenous counterparts, though their presence in the original cell extracts did not interfere with these analyses. The immunoprecipitation permitted the selective isolation of SUMOylation site containing peptides. Moreover, we measured the effect of As<sub>2</sub>O<sub>3</sub> on target modification using label-free quantitative proteomics, and observed an increase in PML SUMOylation in nuclear extracts consistent with previous reports for both SUMO isoforms (Muller, Matunis et al. 1998; Lallemand-Breitenbach, Jeanne et al. 2008). The insertion of an Arg residue at the C-terminus of each SUMO paralog not only facilitated the identification of SUMOylated peptides via the observation of paralog-specific fragment ions, but also reduced the abundance and distribution of overlapping fragment ions observed in MS/MS spectra of peptides with long modified Lys side chain. In addition, the five

amino acid SUMO remnant formed during the tryptic digestion of SUMOylated proteins can also be used for immunoaffinity enrichment. The use of a combined NTA and immunoaffinity enrichment is demonstrated for the identification of SUMO-1 modified peptides.

## 2.3 Experimental section

*Plasmids construction and generation of stable HEK293 cells expressing SUMO constructs.*

cDNA of His6-SUMO wt and His6-SUMO mutants were generated by PCR with the forward primer containing His<sub>6</sub>-tag, KpnI and NcoI restriction sites 5'gaccaagcttggtaccatggctc 3' and the reverse primers containing STOP codon, XhoI restriction site. The WT and SUMO mutant primer sequences were:

*SUMO-1 WT :*

Forward 5' gaccaagcttggtaccatggctc 3'

Reverse 5' ctaccgctcgagtaacccccgttgttctgataaactc 3'

*SUMO-1 mut :*

Forward 5' gaccaagcttggtaccatggctc 3'

Reverse 5' ctaccgctcgagtaacccccgttgttccgataaactc 3'

*SUMO-2 WT :*

Forward 5' gaccaagcttggtaccatggctc 3'

Reverse 5' ctaccgctcgagtaacctcccgtctgctgttgaacacac 3'

*SUMO-2 mut :*

Forward 5' gaccaagcttggtaccatggctc 3'

Reverse 5' ctaccgctcgagtaacctcccgtctgctgtcggaaacacac 3'

*SUMO-3 WT :*

Forward 5' gaccaagcttggtaccatggctc 3'

Reverse 5'ctaccgctcgagttaacctcccgtctgctgctggaacacgctc 3'

*SUMO-3 mut :*

Forward 5' gaccaagcttggtaccatggctcctc 3'

Reverse 5'ctaccgctcgagttaacctcccgtctggttccggaacacgctc 3'

The different His<sub>6</sub>-SUMO constructs were generated by inserting cDNA SUMO paralogs in pCDNA3 or pET28b in KpnI/XhoI or NcoI/XhoI respectively. pcDNA3-PMLIII and pCDNA3-PMLIII-YFP were used as described in Percherancier and *al.* (Percherancier, Germain-Desprez et al. 2009). HEK293 stably expressing SUMO were obtained by transfection with pcDNA SUMO constructs and subsequent neomycine selection (0.5 mg/ml). PML III WT and a PML III 3K construct with mutations at K65, K160 and K490 were obtained as described previously (Percherancier, Germain-Desprez et al. 2009).

*As<sub>2</sub>O<sub>3</sub> treatment and antibodies.*

As<sub>2</sub>O<sub>3</sub> (Sigma) was prepared in 1M NaOH, then further diluted to 1μM in the growth medium, and cells were typically exposed for 4h unless otherwise specified. Rabbit polyclonal anti-SUMO-1 and polyclonal anti-PML antibodies were from Santa-Cruz, rabbit anti-SUMO-2/3 and chicken anti-mouse Alexa-Fluor 594-conjugated secondary antibody from Invitrogen, monoclonal anti- His<sub>6</sub>

antibody from Clontech. HRP-conjugate monoclonal anti-β-actin from Sigma, goat secondary antibodies HRP conjugates from Chemicon International. The TIF-1β antibody was a gift from Dr. M. Aubry (Université de Montréal). Other antibodies used in this study were histone H3 polyclonal antibody (Cell Signaling, cat. #9715), PARP1 antibody (Clontech, cat. # 630210) and the lamin A/C polyclonal (Santa Cruz, #SC-20681).

*Enrichment of SUMOylated proteins*

Cells stably expressing His<sub>6</sub>-SUMO (10<sup>7</sup> cells) were lysed in denaturing buffer A (6 M guanidinium-HCl, 0.1 M NaH<sub>2</sub>PO<sub>4</sub>, 0.01 M Tris-HCl, pH 8, 10 mM β-

mercaptoethanol), sonicated, centrifuged at 16 000g and incubated with 50  $\mu$ L NTA agarose beads (Invitrogen) for 3 h (Tatham, Rodriguez et al. 2009). After washing in buffer A and five times in buffer B (8 M Urea, 0.1 M  $\text{NaH}_2\text{PO}_4$ , 0.01 M Tris-HCl, pH 6.3, 10 mM  $\beta$ -mercaptoethanol), the beads were eluted with 300 mM imidazole in 0.15 M Tris-HCl, pH 6.7, 30 % Glycerol, 0.72 M  $\beta$ -mercaptoethanol. The eluates were subjected to 4-12 % NuPAGE Bis-Tris Gel (Invitrogen).

### *Confocal microscopy*

HEK293 cells, transfected with pcDNA3-His<sub>6</sub>-SUMO and pcDNA3-PMLIII-YFP, were fixed in 4% paraformaldehyde for 10 min at 4°C and revealed by anti-His antibody, followed by Alexa-Fluor 594-conjugated antibody. Confocal images were obtained on a Leica TCS-NT/SP inverted confocal lasers scanning microscope using an Apochromat x 63/1.32 oil-immersion objective. Co-localization experiments were performed by overlaying images using the Leica Confocal Software LCS (Heidelberg, Germany). Excitation and emission filters for the different labelled dyes were as follows: YFP (green):  $\lambda_{\text{ex}}$ : 488 nm,  $\lambda_{\text{em}}$ : 540/25 nm; Texas red (red):  $\lambda_{\text{ex}}$ : 568 nm,  $\lambda_{\text{em}}$ : 610/30 nm. DAPI  $\lambda_{\text{ex}}$ : 405 nm, 10 % power.

### *Recombinant His<sub>6</sub>-SUMO proteins production and purification*

*E. Coli* BL21 cells transformed with different pET28- His<sub>6</sub>-SUMO expressing vectors were induced with 1mM IPTG for 5 h. Cells were lysed in 20 mM phosphate buffer pH 7.6, 500 mM NaCl and 30 mM imidazole by successive liquid nitrogen and 37°C bath, followed by sonication. After centrifugation, the supernatant was loaded on a 5 ml NTA HiTrap Chelating HP column (GE Healthcare). The column was washed according to the manufacturer instructions, and the sample was eluted using an imidazole gradient of 50-500 mM. The fractions containing most of His<sub>6</sub>-SUMO recombinant protein were concentrated on ultra-centricon with a cut-off of 30 kDa (Millipore).

#### *In vitro SUMOylation assay*

To the reaction buffer (20mM  $\text{NH}_4\text{CO}_3$  pH 9, 20 mM NaCl, 0.5 mM DTT), 1  $\mu\text{g}$  recombinant SUMO proteins, 0.5  $\mu\text{g}$  of substrates E2-25K (Boston Biochem), GST-RanGAP fragment 418-587 (Boston Biochem), or GST-PML fragment 485-495 (Biomol International), 0.1  $\mu\text{g}$  SASE1/SAE2 heterodimer (Boston Biochem), 0.5  $\mu\text{g}$  conjugating enzyme hUbc9 (Boston Biochem) were added with or without 5 mM Mg-ATP (Boston Biochem). After incubation at 37°C for 1 h, the reaction was stopped with 10 mM EDTA. The samples were analysed by immunoblot, coomassie staining or silver staining and MS.

#### *Cell fractionation and large scale purification of SUMOylated proteins*

HEK293 cells ( $10^8$  cells/replicate), stably expressing His6-SUMO mutant, were lysed in hypotonic buffer (10 mM Tris-HCl pH 7.65, 1.5 mM  $\text{MgCl}_2$ , 1mM DTT, 20 mM N-Ethylmaleimide, proteases inhibitors) and centrifuged at 3000g. The supernatant constituted the cytoplasmic fraction. The pellet was resuspended in buffer A, sonicated, centrifuged at 16 000g and added to 500  $\mu\text{l}$  NTA agarose beads for 3 h. After washing and elution steps as described above, an aliquot of the eluate was used for immunoblot and the rest (50-60  $\mu\text{g}$ /replicate) for MS analyses. Proteins were reduced in 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) (Pierce) for 20 min at 37°C and then alkylated in 50 mM chloroacetamide (Sigma-Aldrich) for 20 min at 37°C. A solution of 50 mM dithiothreitol was added to the protein solution to react with excess chloroacetamide. Total protein amount was quantitated by Bradford protein assay. Proteins were digested in 50 mM ammonium bicarbonate with modified trypsin (Pierce) overnight (1:30, enzyme:substrate ratio) at 37 °C under high agitation speed. The digest mixture was acidified with trifluoroacetic acid (TFA) and then desalted using a HLB cartridge (Waters) and then dried down by speed vac prior to MS analyses.

#### *Immunoaffinity enrichment of SUMOylated peptides*

NTA-enriched protein extracts of ~50  $\mu\text{g}$ /replicate were reduced with tris(2-carboxyethyl)phosphine (TCEP) (Pierce), alkylated by chloroacetamide (Sigma-Aldrich)

prior to tryptic digestion (1:30 enzyme:substrate, sequencing grade trypsin, Promega, Madison MI). Tryptic peptides were incubated overnight at 4°C with the SUMO-specific monoclonal antibody (10:1, antibody:peptide digest). The antibody-antigen solutions were washed 3 times with 300 µL of TBS buffer and 3 times with 0.1 x TBS in a 10K Microcon centrifugal dialysis tube (Waters, Milford, MA) to remove non-specific binding peptides. SUMOylated peptides were subsequently eluted using 0.2 % formic acid and collected in the non-retained fraction of the centrifugal dialysis unit. Eluates were concentrated on a Speedvac, re-suspended in 3% ACN, 0.2% formic acid and subjected to mass spectrometry analysis.

### *Mass Spectrometry*

All LC-MS/MS analyses were performed using an LTQ-Orbitrap Velos hybrid mass spectrometer with a nanoelectrospray ion source (ThermoFisher, San Jose, CA) coupled with an Eksigent nano-LC 2D pump (Dublin, CA). Peptides were separated on a Optiguard SCX trap column, 5 µm, 300Å, 0.5 ID x 23 mm (Optimize technologies, Oregon City, OR) and eluted on-line to a 360 µm ID x 4 mm, C<sub>18</sub> trap column prior to separation on a 150 µm ID x 10 cm nano-LC column (Jupiter C<sub>18</sub>, 3 µm, 300 Å, Phenomex). Tryptic digests were loaded on the SCX trap and sequentially eluted using salt plugs of 0, 75, 250, 500, 1 and 2M ammonium acetate, pH 3.5. After adsorption on the C<sub>18</sub> precolumn, peptides were separated on the analytical column using a linear gradient of 5-40% acetonitrile (0.2% formic acid) in 53 min and a flow rate of 600 nL/min. The conventional MS spectra (survey scan) were acquired in profile mode at a resolution of 60 000 at  $m/z$  400. MS/MS spectra were acquired in the ion trap using collision-induced dissociation (CID) only or by combining CID and electron transfer dissociation (ETD) with supplemental activation mode in a decision tree data-dependent fashion (Swaney, McAlister et al. 2008) for multiply charged ions exceeding a threshold of 10000 counts. The lock mass option was not used. ETD activation was triggered for precursor ions  $m/z < 650$  (3+),  $m/z < 900$  (4+) and  $m/z < 950$  (5+). For ETD, the precursor cation AGC target was set at 50,000, whereas a value of 100,000 was used for the fluoranthene anion population. Ion/ion reaction duration was fixed at 200 ms. The

dynamic exclusion of previously acquired precursor ions was enabled (repeat count 1, repeat duration: 30 s; exclusion duration 120 s).

Orbitrap raw LC-MS data files were transformed into peptide maps using in-house peptide detection and clustering software (Marcantonio, Trost et al. 2008). Peptide maps belonging to one experiment were clustered and aligned using clustering parameters of  $\Delta m/z = 0.02$  and  $\pm 2$  min (wide),  $\pm 0.5$  min (narrow). Peptide clusters were aligned with mascot identification files to assign sequence identity.

#### *Protein identification and bioinformatic analyses*

MS data were acquired using the Xcalibur software (version 2.0 SR1). Peak lists were then generated using the Mascot distiller software (version 2.1.1, Matrix science) and MS processing was achieved using the LCQ\_plus\_zoom script. Database searches were performed against a nonredundant IPI human database containing 150858 sequences (version 3.54, released Jan 2009) using Mascot (version 2.1, Matrix Science, London, U.K.). A Mascot search against a concatenated target/decoy database consisting of a combined forward and reverse version of the IPI human database was performed to establish a cutoff score threshold of typically 25 for CID or ETD with a false-positive rate of less than 2% ( $p < 0.02$ ). The error window for precursor and fragment ion mass values were set to 0.02 and 0.5 Da, respectively. The number of allowed missed cleavage sites for trypsin was set to 1 and phosphorylation (STY), oxidation (M), deamidation (NQ), carbamidomethylation (C) and SUMOylation (K) (GGTQE: SUMO-1 or GGTQQ: SUMO-2 or GGTQN: SUMO-3) were selected as variable modifications. We also developed a Perl script that searched the mascot generic file (mgf) for the specific SUMO fragment ions (e.g. SUMO-1: m/z 240.1, 258.1, 341.1, 359.2; SUMO-3: m/z 243.1, 344.2; and neutral losses of SUMO remnants) to produce a subset mgf file containing all MS/MS spectra of potential SUMOylated peptide candidates. Potential SUMO peptide candidates were also obtained using ChopNSpice by generating a modified IPI human database containing the five amino acid SUMO-specific tag on each lysine residue (Hsiao, Meulmeester et al. 2009). Database searches using SUMmOn were performed using a custom software available from Dr. Brian Raught

(University Health Network, Toronto, Canada) (Pedrioli, Raught et al. 2006). Manual inspection of all MS/MS spectra for modified peptides was performed to validate assignments.

## 2.4 Results

To facilitate *in vivo* identification of SUMOylated proteins, we developed pcDNA-His-SUMO expression vectors comprising strategically located mutations at the C-terminus of each SUMO paralog (Figure 2.1a). These mutations confer important properties to the stably expressed protein products. First, His<sub>6</sub>-SUMO mutants with an Arg substitution introduce a convenient tryptic cleavage site on the side chain of modified Lys residues whereby individual paralog can be identified by mass-specific signature fragment ions. Second, the short five amino acid segment appended to the modified Lys residues (e.g. EQTGG for SUMO-1 mut, Figure 2.1a) result in fewer fragment ions from the Lys side chain, a property that favor the identification of SUMOylated peptides using conventional database search engines. An additional advantage of the short SUMO remnant is the availability of an epitope to which antibodies can be raised and used in large scale immunoaffinity experiments. In the context of this study, we devised an affinity enrichment strategy whereby SUMOylated proteins are first isolated under denaturing conditions using NTA columns prior to their tryptic digestion and subsequently enriched using immunoaffinity purification. Finally, the SUMO-modified peptides are identified by LC-MS/MS (Figure 2.1b).

*His<sub>6</sub>-SUMO mutants are functional and can be use to monitor protein SUMOylation in vitro and ex vivo*

To determine that site-directed mutagenesis did not impair the transfer of His<sub>6</sub>-SUMO mutants by the SAE1/2-Ubc9 conjugation machinery, we conducted *in vitro* assays using well-established protein SUMOylation substrates. We compared the SUMOylation profiles of His<sub>6</sub>-SUMO-1/2/3 wild type (WT) and the corresponding mutants using RanGAP1 and the ubiquitin-conjugating enzyme E2 (E2-25K), two proteins that are SUMOylated *ex vivo* and *in vitro* (Matunis, Coutavas et al. 1996;

Pichler, Knipscheer et al. 2005). An intact E2-25K and a GST-tagged C-terminal RanGAP1 protein fragment (aa 418-587) were SUMOylated *in vitro* as described previously (Matunis, Coutavas et al. 1996; Rogers, Horvath et al. 2003) (Figure 2.2a). The silver stained gels of the corresponding reactions indicated that all His<sub>6</sub>-SUMO mutants showed conjugation efficiencies comparable to those of WT SUMO proteins. In the presence of ATP, almost all protein substrates were converted to the SUMOylated RanGAP1 and E2-25K products. Interestingly, we observed polySUMOylated chains for all His<sub>6</sub>-SUMO mutants and WT proteins conjugated to RanGAP1, including SUMO-1 that does not contain a  $\psi$ KxE consensus motif. MS analyses of *in vitro* digestion products confirmed the SUMOylation of RanGAP1 for all SUMO paralogs on the residue corresponding to K524 (Annex I, supplementary Figure 2.1). PolySUMOylation was identified on K11 for His<sub>6</sub>-SUMO-2,3 mutants and their WT proteins. We also observed SUMOylation of His<sub>6</sub>-SUMO-1 on sites K23, K37, K39 and K48, two of which (K37 and K39) were previously reported by Cooper et al. during *in vitro* experiments (Cooper, Tatham et al. 2005) (Annex I, supplementary Figure 2.2). *In vitro* SUMOylation analyses confirmed the covalent attachment of His<sub>6</sub>-SUMO-1-3 mutants to K14 of E2-25K (Figure 2.2b). Each His<sub>6</sub>-SUMO mutant was uniquely identified by specific fragment ions ( $b_2^*$ ,  $b_3^*$ ,  $b_2^* - H_2O$ ,  $b_3^* - H_2O$ , etc...) arising from the cleavage of the SUMO side chain for precursor ions of  $m/z < 750$  fragmented in the ion trap (1/3 rule for fragment ion transmission). The correlation of these unique fragment ions using extracted ion chromatograms facilitated the identification of potential SUMOylated peptides in complex tryptic digests. It is noteworthy that all SUMOylated proteins examined, including the fusion protein GST-PML<sub>485-495</sub> were efficiently deSUMOylated by SENP1, thus confirming that the mutation site did not impair the enzymatic activity of the SUMO-isopeptidases (Annex I, supplementary Figure 2.3).

We compared the *ex vivo* SUMOylation efficiencies of His<sub>6</sub>-SUMO mutants with their WT counterparts in HEK293 cells. We first examined the changes in SUMOylation of PML upon treatment with 1  $\mu$ M As<sub>2</sub>O<sub>3</sub> for 4 h. As<sub>2</sub>O<sub>3</sub> is known to enhance the SUMOylation and the subsequent degradation of PML and the PML-RAR $\alpha$  fusion proteins in APL cells (Zhu, Koken et al. 1997; Muller, Matunis et al. 1998; Lallemand-

Breitenbach, Jeanne et al. 2008). Three SUMOylation sites within PML have been reported previously (Kamitani, Kito et al. 1998) though only K160 is required for  $As_2O_3$ -triggered degradation (Zhu, Koken et al. 1997; Lallemand-Breitenbach, Zhu et al. 2001). Immunoblots showed increased polySUMOylation of PML for the His<sub>6</sub>-SUMO WT and mutants upon  $As_2O_3$  treatment (Figure 2.3a, upper panel). The increase in PML SUMOylation was also accompanied by the depletion of the unmodified PML in both WT and mutant SUMO-3. The SUMOylation of PML was clearly evidenced in immunoblots from protein extracts purified using NTA columns (Figure 2.3a, bottom panel). It is noteworthy that PML showed an increase SUMOylation by SUMO-3 and SUMO-2 compared to SUMO-1, a situation that results in ubiquitin-dependent proteolytic degradation of PML (Stefan, Kirstin et al. 2008).

PML is the organizer of subnuclear structures of 0.2-1.0  $\mu$ m named PML nuclear bodies (NBs) that are not only present in most mammalian cell nuclei but also require SUMO for their formation. These subcellular foci are involved in the regulation of different cellular processes, including the induction of apoptosis and cellular senescence, inhibition of proliferation, maintenance of genomic stability and antiviral responses (Bernardi and Pandolfi 2007; Everett and Chelbi-Alix 2007). We examined the recruitment of His<sub>6</sub>-SUMO mutants and WT on NBs in HEK293 cells co-transfected by His<sub>6</sub>-SUMO and YFP-PML construct (Figure 2.3b). Immunofluorescence staining of His<sub>6</sub>-SUMO mutants confirmed their co localization with YFP-PML in multiple and dense nuclear foci characteristics of PML NBs similar to that observed for His<sub>6</sub>-SUMO WT. In addition,  $As_2O_3$  induced aggregation of SUMOylated PML in NBs for both His<sub>6</sub>-SUMO-1 WT and mutant compared to untreated cells (Annex I, supplementary Figure 2.4). Altogether, these experiments established that His<sub>6</sub>-SUMO mutants have functional characteristics similar to those of their WT counterparts.

#### *Subcellular distribution and induction of protein SUMOylation by $As_2O_3$*

To determine the global distribution of SUMOylated proteins, we performed subcellular fractionation to isolate cytosol and nuclear extracts from HEK293 cells

expressing WT and mutant His<sub>6</sub>-SUMO. Immunoblot analyses of these extracts using anti-His antibodies revealed that a higher proportion of SUMOylated proteins was found in nuclear fractions of cells expressing His<sub>6</sub>-SUMO-1 and His<sub>6</sub>-SUMO-3 mutants (Figure 2.4a, lanes 4 and 6). While polySUMOylation chains were observed for high molecular weight bands of these two paralogs, higher polymerization levels were noted for proteins modified with His<sub>6</sub>-SUMO-3 consistent with previous reports (Saitoh and Hinchee 2000). Interestingly, free His<sub>6</sub>-SUMO-1 and His<sub>6</sub>-SUMO-3 were more abundant in the cytosol compared to nuclear extracts (Figure 2.4a, lanes 3 and 5) as previously noted by Seeler et al (Seeler and Dejean 2003). It is noteworthy that anti-His immunoblots also revealed the presence of non-specific proteins in nuclear extracts of mock HEK293 cells (Figure 2.4a, lanes 1 and 2). MS analyses of these NTA-purified nuclear extracts (see below) identified several non-specific proteins including Forkhead box and homeobox proteins, POU domain transcription factors, Histidine triad nucleotide-binding proteins, that contain multi-His sequences and Zn metal-binding proteins known to bind to Ni<sup>2+</sup> ions (Annex IV, CD-ROM supplementary Table I and 2.2).

Overall changes in protein SUMOylation were also evaluated in NTA-purified nuclear extracts from cells treated or not with As<sub>2</sub>O<sub>3</sub> (Figure 2.4b). Enhanced protein SUMOylation was noted for both His<sub>6</sub>-SUMO-1 and His<sub>6</sub>-SUMO-3 mutants resulting in multiple band patterns of high molecular weight proteins for the corresponding SUMO-1 and SUMO-2/3 immunoblots. It is noteworthy that multimerization can be obtained with mixed SUMO chains from endogenous and mutant proteins, the distribution of which depends on their relative proportion *ex vivo*. Interestingly, PML immunoblots of the same NTA-affinity purified extracts clearly showed an increased in the SUMOylation of endogenous PML upon As<sub>2</sub>O<sub>3</sub> treatment (Figure 2.4b, upper panel). The corresponding banding pattern is almost superimposable to that of both SUMO-1 and SUMO-2/3 immunoblots, suggesting that PML represents a primary SUMOylation substrate upon As<sub>2</sub>O<sub>3</sub> treatment (Figure 2.4b, bottom panel). Note that in the absence of As<sub>2</sub>O<sub>3</sub>, endogenous PML is barely detectable even when using longer exposure periods (data not shown). Although we have identified TIF-1β, Lamin A/C, Poly (ADP-ribose)

polymerase 1 (PARP1) and Histone H3 as modified proteins by SUMO (see below), we could not detect the SUMOylation of these endogenous proteins by immunoblots, presumably due to the low abundance of their SUMOylated counterparts in cell extracts and/or the inaccessibility of epitope for antibody binding (data not shown). Among all the proteins tested, only the endogenous PML showed enhanced SUMOylation in response to  $As_2O_3$  (Figure 2.4b, bottom panel).

#### *Large scale identification of protein SUMOylation*

To identify SUMOylated proteins present in nuclear extracts, we performed large scale NTA-affinity purification experiments from HEK293 cells expressing His<sub>6</sub>-SUMO-3 mutant exposed or not to  $As_2O_3$ . Similar experiments were also performed on mock HEK293 cells to identify proteins binding non-specifically to the NTA affinity column. We typically obtained 40-60  $\mu$ g of NTA-purified proteins from  $10^8$  HEK293 cells in any of the conditions and biological replicates examined. Proteins extracts following NTA purification (2  $\mu$ g/injection) were subjected to MS analyses using a nano 2D-LC system (SCX/C<sub>18</sub>) coupled to a LTQ-Orbitrap Velos instrument. Tandem mass spectra were acquired using CID and ETD in a decision tree manner to enhance the overall number of identification (Swaney, McAlister et al. 2008). In total, we acquired more than 15,000 MS/MS spectra corresponding to 6282 unique peptides identified using Mascot database search engine. To reduce the number of ambiguous identification, we compared proteins that were identified by at least 2 peptides in each condition with a FDR of less than 2 %. By using these conservative selection criteria, we identified a total of 639 unique proteins, of which 232 proteins (36%) were common to all three different cell extracts (Figure 2.5a). Common proteins were assigned to non-specific binders co-purified from NTA columns, and included proteins containing multi-His sequences and Zn metal-binding proteins. It is noteworthy that some of these proteins such as E3 SUMO-protein ligase RanBP2 (RanBP2), Zinc finger protein OZF, DNA topoisomerase 1 (Top1) were previously reported to be SUMOylated, and their fortuitous isolation from mock HEK293 NTA extract could not definitively rule out their endogenous SUMOylation. More importantly, we identified 205 proteins specific to HEK293 SUMO-3 mutant extracts (with and without  $As_2O_3$ ) that comprised known SUMOylated

substrates such as PML, Transcription Intermediary Factor 1-beta (TIF-1 $\beta$ ), PARP1, heterogeneous nuclear ribonucleoprotein isoform F (hnRNP F) and hnRNP C1/C2. A list of proteins found in protein extracts from each condition is provided in Annex IV CD-ROM supplementary Table I. It should be noted that unambiguous assignment of SUMOylated proteins relies on the identification of tryptic peptides comprising the five amino acid SUMO remnant on the modified Lys residues, a situation that only applies to a smaller subset of the potential SUMOylated proteins (see below). We also performed Gene Ontology analysis for terms associated with biological processes enriched in the HEK293 SUMO-1/3 mutant extracts using the software Protein ANalysis THrough Evolutionary Relationships, PANTHER (<http://www.pantherdb.org/>). These analyses revealed that potential SUMOylated targets were significantly enriched in proteins involved in chromatin remodeling, organelle organization, and nuclear transport (Supplementary Figure 2.5).

The comparison of NTA-enriched protein extracts from cell expressing His<sub>6</sub>-tag SUMO-3 mutant enabled the identification of at least 205 SUMOylated protein candidates that were not detected in mock HEK293 cells. To identify the location of SUMOylation sites on protein substrates, we used Mascot, SUMmOn and ChopNSpice search engines. We also developed a script to make use of the specific SUMO fragment ions in order to retrieve all MS/MS spectra of potential SUMOylated peptide candidates (Experimental procedures). Altogether, we identified 17 unique SUMOylation sites on 12 different protein substrates from these large-scale proteomics experiments (Table I). All MS/MS spectra were validated manually and comprised fragments characteristic of the SUMO-3 side chain (GGTQN). A distribution of the number of identified residues according to the three different database search engines is provided in Annex IV CD-ROM Supplementary Table III. We confirmed previously known SUMOylation sites on proteins such as PML (K490) and TIF-1 $\beta$  (K750, K779) and cross-link sites on SUMO-2/3 (K11, K41 for SUMO-3 and K42 for SUMO-2). These analyses also revealed new SUMOylation sites on previously unreported nuclear substrates such as histone H3 (K23), Lamin (K420), SAFB2 (K524), Rsf1 (K287), WIZ1 (K1523), and cross-link sites with SUMO-4 (K11) and ubiquitin (K11).

To profile proteins that showed differential regulation upon cell treatment with  $\text{As}_2\text{O}_3$ , we compared the abundance of peptide ions identified in digests of HEK293 cells expressing His<sub>6</sub>-SUMO-3 mutant from control and stimulated cells. The distribution of abundance for 6790 peptide ions is shown in the scatter plot of Figure 2.5b, and indicated that more than 92 % of all ions showed less than 3-fold change in abundance upon cell stimulation. The most significant change was observed for PML, a protein that showed more than 15-fold increase upon  $\text{As}_2\text{O}_3$  treatment. We obtained a sequence coverage of 43 % for this protein, and several PML tryptic peptides were found to be modified with SUMO-3 mutant. For example, the product ion of the quadruply-protonated peptide ion ( $m/z$  523.5) acquired using ETD fragmentation is shown in Figure 2.5c and confirmed the SUMOylation of the K490 residue. The MS/MS spectrum is dominated by c- and z-type fragment ions from the peptide backbone and by specific fragment ions arising from the cleavage of the SUMO-3 mutant side chain (e.g.  $c_2^*$ ,  $c_3^*$ ,  $c_4^*$ ). As indicated in Figure 2.5b, the abundance of this peptide was increased in samples from cells treated with  $\text{As}_2\text{O}_3$ . Residue K490 is one of three known PML SUMOylation sites, the other two being K65 and K160 (Kamitani, Kito et al. 1998). However, we could not identify tryptic peptides harboring these two modified residues in any of the cell extracts examined, presumably due to the relatively large molecular weight and hydrophobicity of the corresponding peptides that precluded their successful separation by  $\text{C}_{18}$  chromatography.

Our MS analyses also revealed the presence of three new PML SUMOylation sites at K380, K400 and K497. Residues K380 and K400 were previously identified as sites of polyubiquitylation in response to  $\text{As}_2\text{O}_3$  (Lallemand-Breitenbach, Jeanne et al. 2008). Site-directed mutagenesis indicated that mutation of K400 delayed but did not prevent PML ubiquitylation and its subsequent proteasome-mediated degradation (Lallemand-Breitenbach, Jeanne et al. 2008). Residues K380 and K400 are located between the B box domains and the nuclear localization sequence (NLS) whereas K497 is next to the NLS of PML (Figure 2.6a). To confirm the identification of these new SUMOylation sites, we first examined possible site-specificities of different SUMO paralogs with

transfected PML and SUMOylation-site mutants thereof. We compared the *ex vivo* SUMOylation efficiency by each SUMO WT paralog in PMLIII WT and a PMLIII 3K mutant (K65R, K160R and K490R) in extracts from HEK293 cells co-transfected with the different PML and SUMO constructs (Figure 2.6b). Anti-PML immunoblots showed an increase in PMLIII WT SUMOylation by SUMO-2 and SUMO-3 when cells were exposed to As<sub>2</sub>O<sub>3</sub>, consistent with results presented in Figure 2.3a. This was clearly evidenced for the same protein extracts purified using NTA columns (Figure 2.6b, His pull down). It is noteworthy that similar experiments performed using more sensitive ECL immunoblots revealed the SUMOylation of PML by SUMO-1, but to a lower level than that observed for SUMO-2 and SUMO-3 (Annex I, supplementary Figure 2.6c). In all cases, PML showed significantly higher SUMOylation levels by SUMO-3 compared to other SUMO paralogs. In contrast, PMLIII 3K displayed one band in immunoblot analysis of extracts from cells co-transfected with PMLIII 3K and His<sub>6</sub>-SUMO WT treated or not with As<sub>2</sub>O<sub>3</sub> (see input Figure 2.6b, upper right panel). Interestingly, NTA protein extracts of these samples indicated that residual SUMOylation of PMLIII 3K was observed with His-SUMO-3, and that this modification was enhanced upon As<sub>2</sub>O<sub>3</sub> treatment (see His pull down Figure 2.6b, middle right panel). These experiments indicated that in the absence of the three known SUMOylation sites (K65, K160 and K490), PML is still SUMOylated, though the extent of this modification is significantly lower than that observed for PMLIII WT. The new sites identified could thus account for the remaining SUMOylation observed in PML.

LC-MS/MS analyses of tryptic digests from NTA-purified protein extracts of HEK293 His<sub>6</sub>-SUMO-3 mutant cells identified three additional PML SUMOylation sites modulated by As<sub>2</sub>O<sub>3</sub>. The extracted ion chromatograms of multiply-charged ions corresponding to modified tryptic peptides at residues K490, K400 and K380 are presented in Figure 2.6c and indicated that all three peptides are up-regulated by at least 10-fold upon cell stimulation with As<sub>2</sub>O<sub>3</sub>. The tryptic peptide with SUMOylated K490 and K497 residues was also increased in abundance upon As<sub>2</sub>O<sub>3</sub> treatment (data not shown). The MS/MS spectra of precursor ions  $m/z$  392.54<sup>3+</sup> and  $m/z$  835.89<sup>2+</sup> showed abundant fragment ions from which the location of the modified residues could be

clearly established. It is noteworthy that the modified K380 and K400 residues are not harboring the consensus motif  $\psi$ KxE/D. We also validated these assignments using synthetic peptides bearing the pentapeptide SUMO stub at the modified Lys residues. The MS/MS spectra of the corresponding peptides are shown as supplementary data (annex II and III) and yielded fragmentation patterns superimposable to those of the native peptides, thereby confirming two previously unknown SUMOylation sites on PML (Annex I, supplementary Figure 2.7).

*Identification of SUMOylated peptides using a dual affinity enrichment approach*

The expression of His<sub>6</sub>-SUMO mutants in HEK293 cells enabled a convenient approach to identify SUMOylated proteins and to distinguish them from proteins binding in a non-specific manner to the NTA column. The identification of modified residues in MS/MS experiments is also facilitated by the observation of fragment ions characteristic of SUMO remnants. While this approach enabled the identification of modified residues in a small subset representing less than 0.5 % of the entire peptide population, the detection of protein SUMOylation for paralogs expressed at a lower abundance (e.g. SUMO-1) remained a challenging task. Indeed, we could not identify any SUMO-1-modified tryptic peptides when similar 2DLC-MS/MS experiments were performed on NTA-enriched protein extracts from HEK293 His<sub>6</sub>-SUMO-1 mutant cells (data not shown).

To facilitate the identification of SUMOylated residues from NTA protein extracts of HEK293 His<sub>6</sub>-SUMO-1 mutant cells, we raised monoclonal antibodies against the SUMO-1-specific GGTQE epitope. We optimized the binding and elution protocol using synthetic SUMOylated peptides from PML (K490) and E2 ligase (K14) spiked at different levels into HEK293 tryptic digests. Immunoaffinity binding experiments were performed overnight at 4°C using a molar ratio of 10:1 antibody:antigen. Non-binding peptides were washed with TBS in a 10K centrifugal dialysis tube, and SUMOylated peptides were subsequently eluted using 0.2 % formic acid (Annex I, supplementary Figure 2.8a). Under optimized conditions, we obtained recovery yield of 30-70 % for

peptides spiked at 100-500 fmoles in a 9 ug tryptic digest of HEK293 (Annex I, supplementary Figure 2.8b).

This dual affinity purification approach was applied to cell extracts for mock HEK293 and HEK293 His<sub>6</sub>-SUMO-1 mutant following cell stimulation with As<sub>2</sub>O<sub>3</sub>. In total, we identified 215 proteins unique to SUMO-1 extract in all three different extracts including transcription factors (Transcriptional repressor protein YY1, PML), DNA and RNA binding proteins (H2A, SAFB2). The complete protein list can be found in Annex IV, Supplementary Table II. The 2D-LC-MS/MS analysis of the tryptic digest following immunoaffinity purification enabled the identification of 4 SUMOylated tryptic peptides unique to the HEK293 His<sub>6</sub>-SUMO-1 mutant extract. It is noteworthy that approximately 48 % of the proteins identified following the immunoaffinity purification were assigned to non-specific binders since they were also observed in the mock HEK293 samples though their abundances varied significantly across sample sets. Our analyses identified 4 tryptic peptides bearing the remnant SUMO-1 at the modified lysine residue, including peptides from PML (K490), RanGAP1 (K524) and SAFB2 (K294) (see Table II). A comparison of the contour map of m/z vs. time is shown in Figure 2.7a for a tryptic digest of a NTA-enriched protein extract before and after immunoaffinity enrichment. The contour map shows an expanded region of the PML tryptic peptide bearing the modified K490 residue. The triply-protonated peptide ion at m/z 702.7 was hardly detectable in the NTA-enriched tryptic digest (Figure 2.7a, left panel) and was not sequenced by data dependent acquisition due to its low abundance (< 20000 counts). However, the same peptide ion was clearly evidenced following the dual affinity purification (Figure 2.7a, right panel), and yielded an MS/MS spectrum that was successfully assigned to corresponding modified peptide (see below).

To profile proteins that showed differential SUMOylation upon cell treatment with As<sub>2</sub>O<sub>3</sub>, we compared the abundance of peptide ions following the dual affinity purification of digests from HEK293 cells expressing His-SUMO-1 mutant from control and stimulated cells. We monitored the abundance of 836 peptide ions identified in replicate analyses of HEK293 His SUMO-1 mutant cells with and without stimulation

with As<sub>2</sub>O<sub>3</sub> (Annex I, supplementary Figure 2.9). Examples of extracted ion chromatograms are shown in Figure 2.7b for modified tryptic peptides of PML (K490), and RanGAP1 (K524). As observed, the PML tryptic peptide bearing the SUMOylated K490 residue was only detected in cell extracts exposed to As<sub>2</sub>O<sub>3</sub>. This result is consistent with immunoblots shown in Figure 2.4b confirming the increased modification of PML by SUMO-1 upon As<sub>2</sub>O<sub>3</sub> treatment. In contrast, RanGAP1, a well known SUMOylation target, did not show any significant changes in SUMOylation under the same conditions (Figure 2.7b, right panel). Confirmation of the sites of modification was obtained from the product ions spectra of the corresponding precursor ions (Figure 2.7c). All SUMOylated residues including SAFB2 (K294) displayed the consensus motif  $\psi$ KxE/D.

## 2.5 Discussion

The identification of SUMOylation sites on protein substrates represents an important analytical challenge owing to the relatively low stoichiometry and the dynamic nature of this modification. In mammalian cells, this difficulty is further exacerbated by the large peptide remnant of SUMO paralogs left on the modified lysine residue upon tryptic digestion. In this context, the construction of SUMO paralogs that comprise a His tag and a strategically located Arg residue on the C-terminus provide a convenient approach to enrich and identify short peptides bearing the SUMO-specific remnant moieties. The judicious location of a C-terminus Arg residue was important in order to maintain the function of SUMO paralogs while minimizing structural changes compared to the endogenous proteins. Accordingly, plasmids were constructed with single base substitutions at the C-terminal end of SUMO paralogs to generate mature proteins with an Arg residue at the 6<sup>th</sup> position from the C-terminus, similar to that found in the yeast Smt3 protein.

*In vitro* SUMOylation assays with well known protein substrates such as RanGAP1 and E2-25K confirmed the functionality of the His<sub>6</sub>-SUMO mutants and their efficient transfer by the SAE1/2-Ubc9 conjugation machinery. MS analyses of the enzymatic products not only determined the expected modification sites on these substrates, but also identified sites of polySUMOylation for each paralog. In addition to the site K11 previously reported on SUMO-2 and SUMO-3 proteins, we also identified polySUMOylation sites on SUMO-1 at K23, K37, K39 and K48, though the physiological relevance of SUMO-1 polymerization chains is presently unknown. Indeed, SUMO-1 which lacks the consensus motif  $\psi$ KxE/D has not been reported to form polymeric chains although recent studies indicated that it can cap poly-SUMO-2,3 chains (Martin, Wilkinson et al. 2007). Our analyses also indicated that MS/MS spectra of modified peptides afforded backbone sequence ions together with short SUMO-specific fragment ions from CID and ETD activation that can be advantageously exploited to confirm SUMOylated peptides. The presence of paralog-specific fragment ions such as m/z 240.1 ( $b_2^* - H_2O$ ), 341. 2 ( $b_3^* - H_2O$ ) for SUMO-1 mutant and m/z 243.1

(b<sub>2</sub>\*), 343.2 (b<sub>3</sub>\*) for SUMO-3 mutant can be used for confirmation purposes as shown here or for targeted identification of these modified peptides via data-dependent acquisition. The relatively limited number of fragment ions originating from the modified Lys side chain also facilitated protein identification using common database search engines such as Mascot (Annex IV, Supplementary Table III).

The subcellular localization of SUMOylated proteins using immunoblotting and immunofluorescence experiments revealed that a large proportion of substrates are nuclear, an observation that also accounts for the significant role of this modification in transcription, DNA repair, nuclear bodies and nucleocytoplasmic transport. This distribution is partly attributed to the enrichment of SUMO-modifying enzymes in this compartment, although a sizable number of substrates are also present in the cytoplasm, plasma membrane, mitochondria and endoplasmic reticulum (Geiss-Friedlander and Melchior 2007). We performed large-scale proteomics analyses of nuclear protein extracts from mock and His<sub>6</sub>-SUMO-3 mutants HEK293 cells to identify the nature of SUMOylated substrates including those that could be regulated by As<sub>2</sub>O<sub>3</sub>. By using strict comparison criteria, we found more than 205 proteins unique to the His<sub>6</sub>-SUMO-3 mutants HEK293 cells such as proteins involved in chromatin remodeling, organelle organization, and nuclear transport (Annex I, supplementary Figure 2.5). Following immunoprecipitation, 215 unique proteins were identified in SUMO-1 cells that were absent in the mock. Interestingly, we found several proteins involved in the regulation of ribosome biogenesis including hnRNP proteins, RNA helicases, and ribosomal subunits, suggesting that SUMO-3 modification may regulate the assembly of these macromolecular complexes. Recent reports indicated that several of these substrates were identified in nucleolus extracts and appeared to be regulated through the ubiquitin-proteasome pathway, suggesting that SUMOylation may target unassembled ribosomal proteins for degradation (Matafora, D'Amato et al. 2009) (Lam, Lamond et al. 2007).

Our proteomics analyses also enabled the identification of several new SUMOylation sites in proteins such as histone H3 (K23), Lamin (K420), SAFB2 (K524), Rsf1 (K287), WIZ1 (K1523), and cross-link sites with SUMO-4 (K11) and

ubiquitin (K11). Several of the modified proteins are involved in transcription such as TIF-1 $\beta$ , HSF4B and PML. Of particular interest is PML, a protein that localizes to NBs where it also acts as a tumor suppressor through the regulation of p53 response to oncogenic signals (Pampin, Simonin et al. 2006). Quantitative proteomics revealed that PML showed more than 15-fold increase in abundance upon cell stimulation with As<sub>2</sub>O<sub>3</sub>. In response to As<sub>2</sub>O<sub>3</sub>, PML is phosphorylated through the mitogen-activated protein kinase (MAPK) pathway leading to its transfer from the nucleoplasm to the nuclear matrix, and to an increase in PML SUMOylation and NBs size (Lallemand-Breitenbach, Zhu et al. 2001; Hayakawa and Privalsky 2004). SUMOylated PML recruits the RING-domain-containing ubiquitin E3 ligase, RNF4, resulting in its degradation through the ubiquitin-proteasome pathway.

Interestingly, a PMLIII 3K where all three known sites of SUMOylation were mutated to Arg is still transferred to the nuclear matrix but is resistant to As<sub>2</sub>O<sub>3</sub>-induced PML degradation (Lallemand-Breitenbach, Zhu et al. 2001). The exact mechanism by which PML is transferred to the nuclear matrix in a SUMO-independent manner upon As<sub>2</sub>O<sub>3</sub> treatment is still unclear, but could involve its prior phosphorylation. It is noteworthy that the SUMOylated forms of PML were barely detectable when total extracts from control and As<sub>2</sub>O<sub>3</sub>-treated cells expressing PMLIII 3K and SUMO paralogs were analyzed by immunoblot with anti-PML antibody ((Lallemand-Breitenbach, Zhu et al. 2001) and Figure 2.6b). Furthermore, we observed that NTA enrichment of protein extracts revealed residual SUMOylation of PMLIII 3K by SUMO-3 and that SUMOylation of PMLIII 3K by SUMO increased in response to As<sub>2</sub>O<sub>3</sub>. Our data demonstrated that As<sub>2</sub>O<sub>3</sub>-mediated SUMOylation of PMLIII can still occur at sites other than the three known residues K65, K160 and K490. Detailed proteomics analyses enabled the unambiguous identification of K380, K400, and K497 as additional SUMOylation sites regulated by As<sub>2</sub>O<sub>3</sub> treatment. Interestingly, two of these sites (K380 and K400) were previously shown to be ubiquitylated *in vitro* (Tatham, Geoffroy et al. 2008). Further investigations are required to determine the significance of these new sites on PML functions.

The insertion of a tryptic site close to the C-terminus of SUMO facilitates the MS/MS interpretation but also renders possible the immunoenrichment of SUMO modified peptides. Anti EQTGG (SUMO-1) monoclonal antibody was generated to purify SUMOylated peptides. Prior to the IP, the samples are desalted, since the antibody/antigen reaction is performed in native conditions. Samples were analyzed before (data not shown) and after the IP. No SUMOylation sites were identified prior to the second enrichment for SUMO-1. Detailed analysis of raw data, suggested that SUMO peptides are present in the sample prior to enrichment, but the presence of contaminant peptides impair their identification. DDA favors the sequencing of high abundance species, and SUMOylated peptides that are of significantly lower abundance are under represented in the population of sequenced peptides as illustrated in Figure 2.7. The application of dual affinity purification enables the enrichment and identification of low abundance SUMO-1 modified peptides. We monitored the effect of  $As_2O_3$  on SUMOylated proteins from HEK293 cells and observed the differential SUMOylation of PML K490 while RanGap K524 SUMOylation remained unaffected. Our mass spectrometry data suggest that  $As_2O_3$  treatment primarily affect the SUMOylation of PML, an observation that was correlated by western blot data.

The availability of functional SUMO mutants that can be stably express in human cells opens up new avenues for large-scale SUMOylome analyses. The enrichment of SUMOylated peptides can be achieved using single or dual affinity purification with NTA column or in combination with immunoaffinity to enriched SUMOylated peptides bearing the SUMO remnant motif. We anticipate that the combination of a dual affinity enrichment approach will yield a larger proportion of SUMOylated tryptic peptides than that achievable with NTA alone for SUMO-3. The presence of short paralog-specific peptide segment on the side chain of modified Lys residues provides characteristic fragment ions that facilitate the identification and confirmation of SUMOylated peptides from complex cell digests. These fragment ions can be used advantageously in the design of MS data-dependant experiments to target their identification more efficiently. While MS/MS spectra of these modified tryptic peptides contain spectral features specific to the SUMO remnants, database search of the corresponding peptides can be

efficiently achieved using conventional search engines like Mascot. The analytical advantages of the present approach and the possibility of conducting quantitative proteomics analyses will greatly facilitate large-scale experiments to unveil the complex regulation of protein SUMOylation.

## 2.6 Acknowledgements

The authors thank Dr. Brian Raught and Tharan Srikumar (UHN, Toronto, Canada) for their valuable assistance in the identification of SUMOylated substrates using the software SUMmOn. We are most grateful to Dr. Muriel Aubry (UdeM, Montreal, Canada) for valuable comments and discussions. This work was carried out with the financial support of operating grants from the National Science and Engineering Research Council (NSERC), the Canadian Institutes for Health Research (CIHR), the Canada Research Chair program (PT, SM) and the Ligue Nationale contre le Cancer. IRIC is supported in part by the Canadian Center of Excellence in Commercialization and Research, the Canadian Foundation for Innovation and the Fonds de la Recherche en Santé du Québec (FRSQ). LM is a recipient of a graduate student scholarship from the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT).

**Table I : List of confirmed SUMOylated peptides from NTA-enriched of nuclear proteins from HEK293 His<sub>6</sub>-SUMO-3 mutant**

<b>Proteins</b>	<b>Site*</b>	<b>Status<sup>#</sup></b>	<b>Function</b>
Histone 3.3	K24	Unknown	Nucleosome assembly
Histone 4	K13	Unknown	Nucleosome assembly
HSF4B	K288	Unknown	DNA-binding; Binds HSE; regulates transcription
Lamin A/C	K420	Unknown	Nuclear lamina
PML	K380,K400,K490,K497	Unknown/unknown/known/unknown (Kamitani, Kito et al. 1998)	Transcription factor
RSF1	K287	Unknown	Assembly of nucleosome by RSF chromatin-remodeling complex
SAFB2	K524	Unknown	Binds to scaffold/matrix attachment region. Inhibit cell proliferation
TRIM28	K750, K779	Known (Masclé, Germain-Desprez et al. 2007)	Forms a complex with a KRAB-domain TF and ↑ KRAB-mediated repression
WIZ1		Unknown	Link EHMT1/2 to CTBP corepressor machinery

Ubiquitin	K1523	Unknown	Cross-link
SUMO-2,3	K11	Known (Cooper, Tatham et al. 2005)	Cross-link
	K42		Cross-link
SUMO-4	K11	Expected	Cross-link
	K11	Known (Tatham, Jaffray et al. 2001)	

\*Sites refer to protein sequence including initiating Met residue.

#Status indicates whether or not identified sites were previously reported along with the corresponding reference.

**Table II :List of confirmed SUMOylated peptides from the immunoprecipitation of nuclear proteins from HEK293 His6-SUMO-1 mutant**

Proteins	Site*	Status <sup>#</sup>	Function
RanGap	K524	Known	GTPase activator of Ran protein.
PML	K490	Known	Probable transcription factor.
SAFB2	K294	Known	Binds to the S/MAR of DNA. Inhibit cell proliferation.

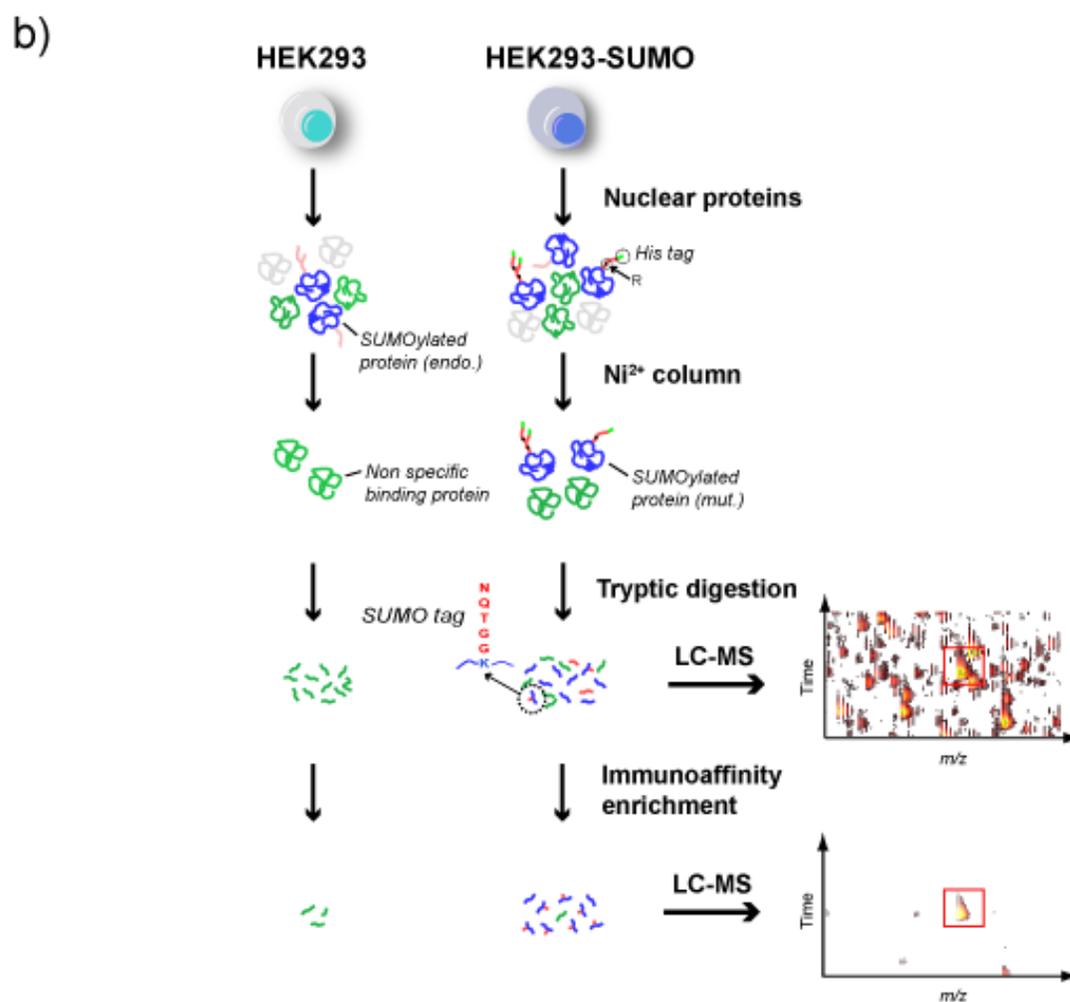
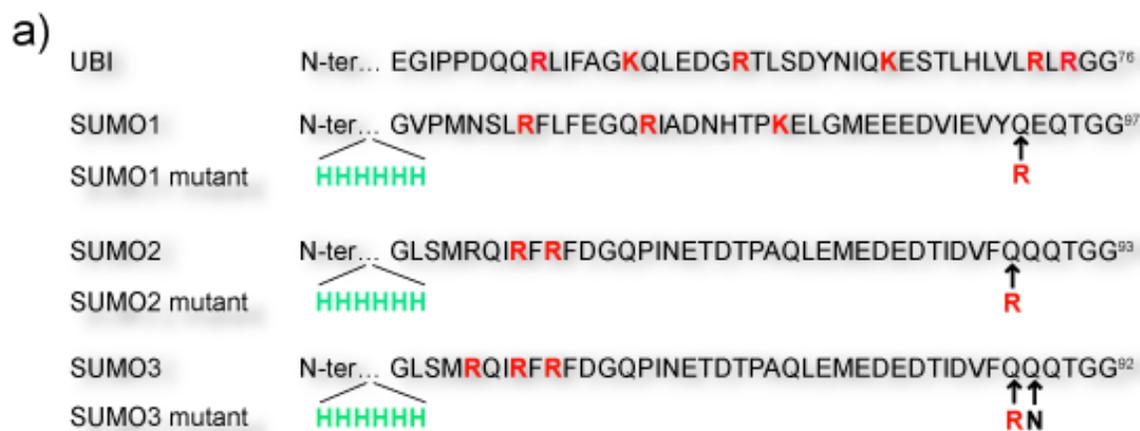
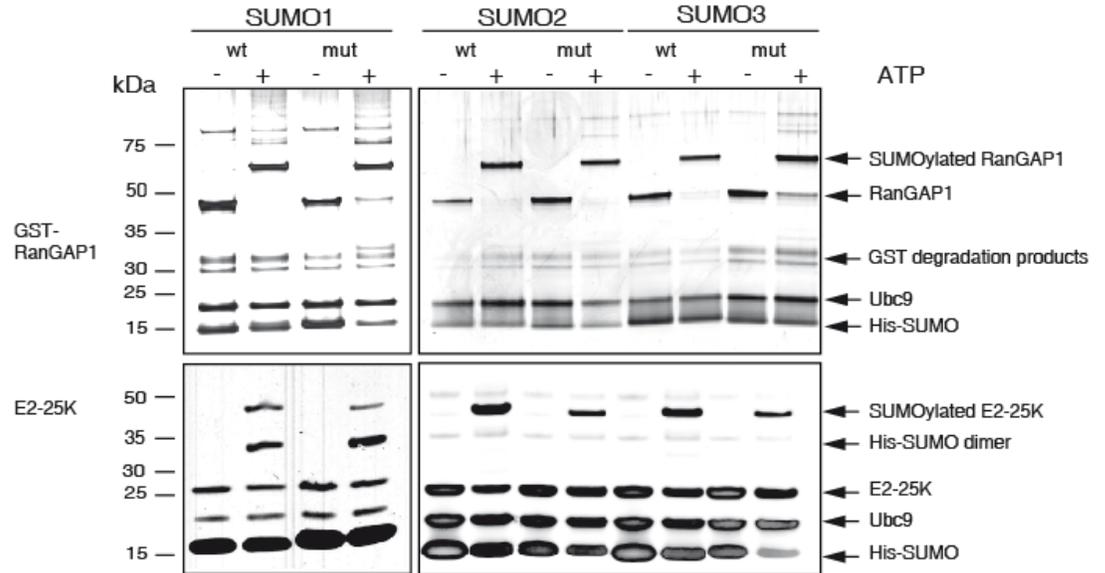


Figure 2.1: Overview of proteomics approach to the identification of SUMOylated peptides.

a)



b)

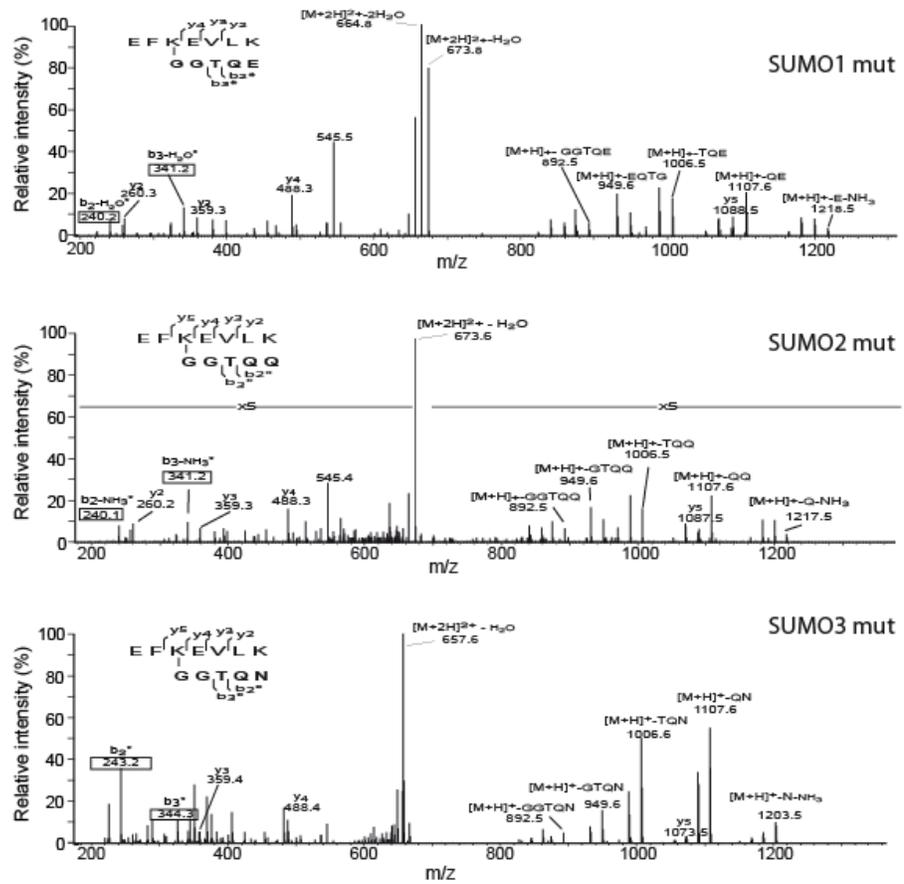


Figure 2.2: In vitro SUMOylation assays of a) RanGAP (K524) and E2-ligase (K14). b) MS/MS confirmation of modified K14 in E2-25K for individual SUMO paralogs.

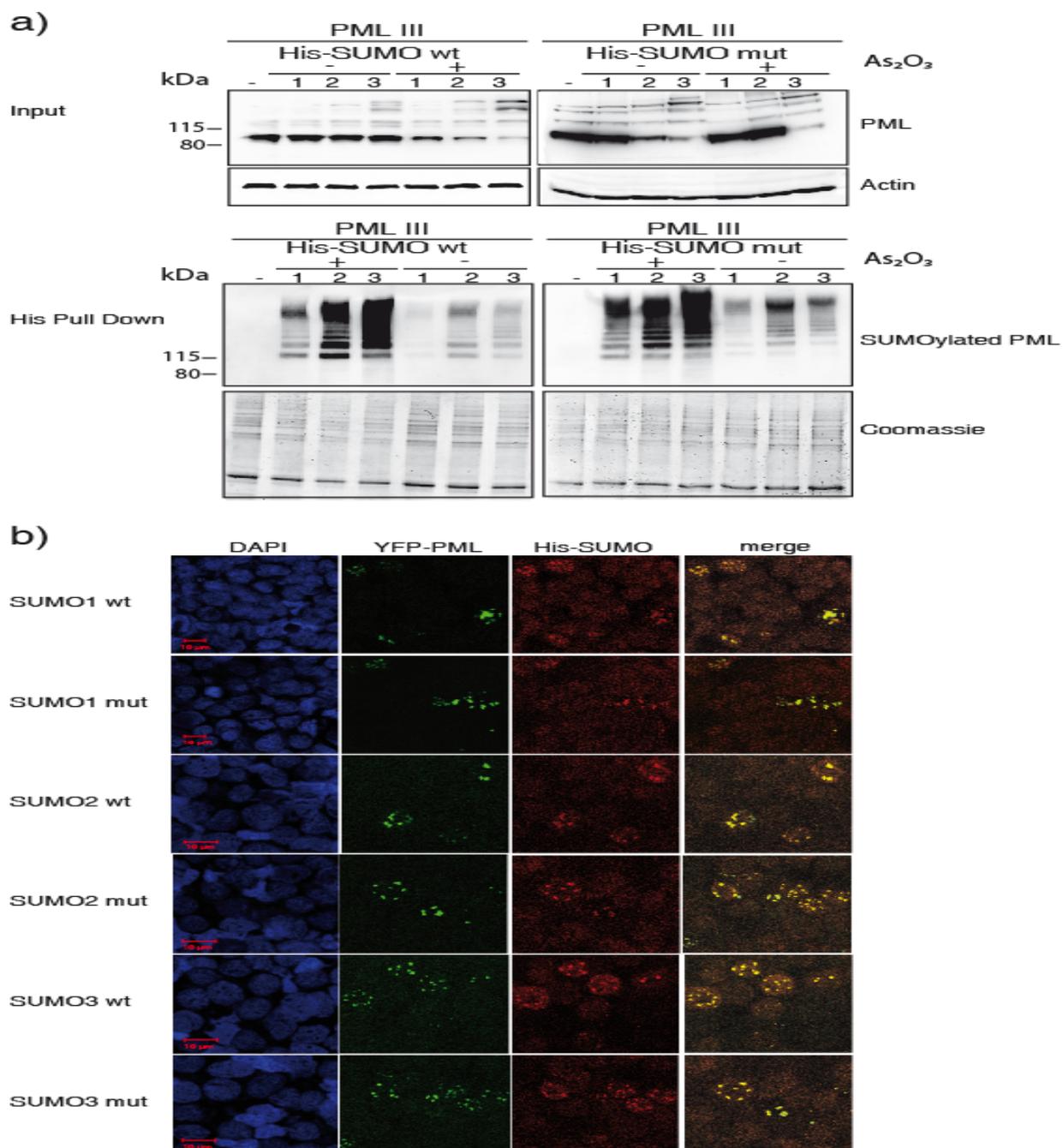


Figure 2.3 : Comparison of His-SUMO WT and mutants to SUMOylate PML and to colocalize within nuclear bodies. a) Immunoblots of input (upper panel) or His pull down (lower panel) of extracts from HEK293 cells, co-transfected with PML and WT or mutant SUMO, treated or not with  $As_2O_3$ . b) Immunofluorescence of HEK293 cells co-transfected with YFP-PML and His-SUMO WT or mutant revealed using anti-His antibody.

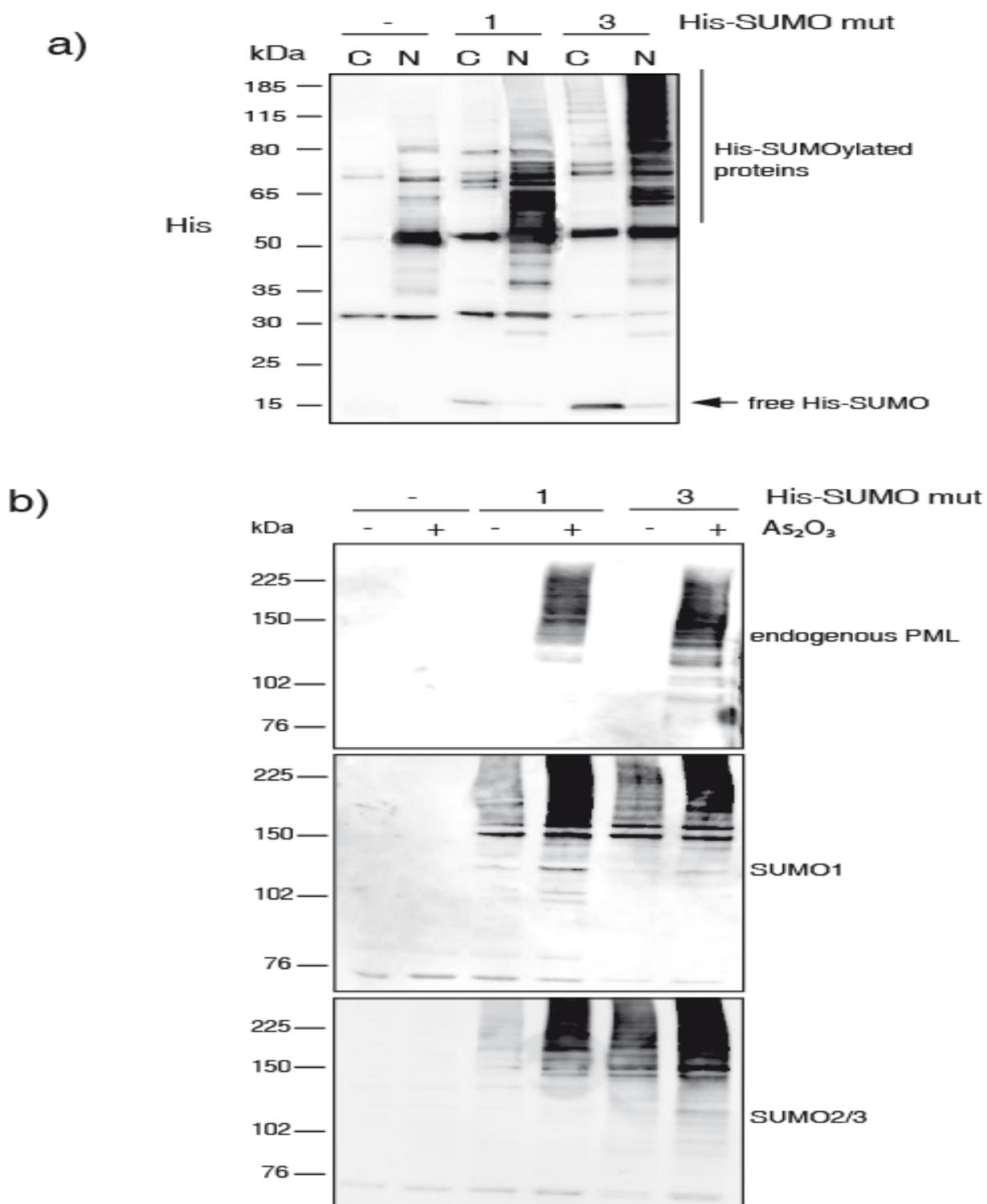


Figure 2.4: Immunoblots of NTA purified extracts from control HEK293 and HEK293-His-SUMO cells. a) Comparison of cytosol (C) and nuclear (N) extracts using anti-His antibody. b) Increased protein SUMOylation in response to As<sub>2</sub>O<sub>3</sub> revealed using anti-PML, anti-SUMO-1 and anti-SUMO-2/3 antibodies.

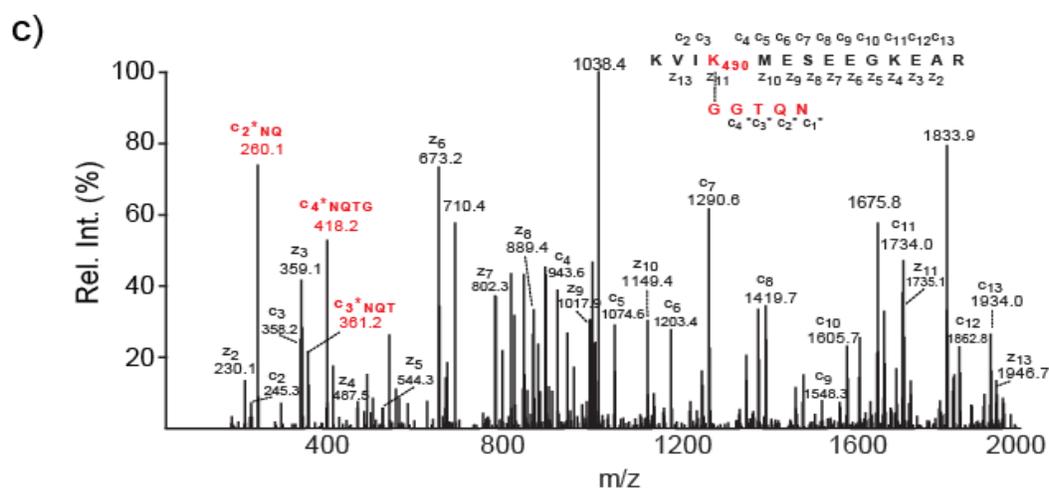
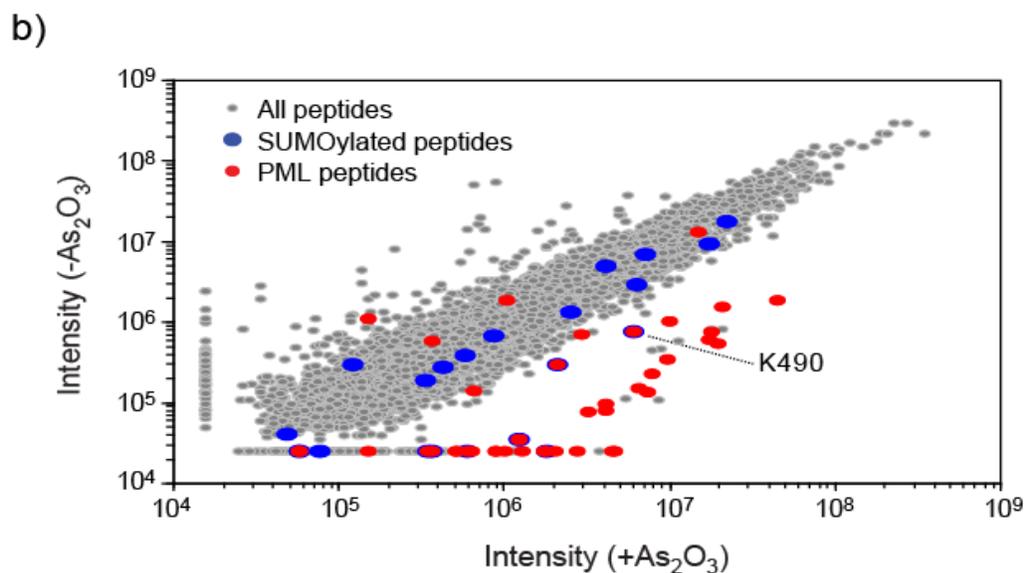
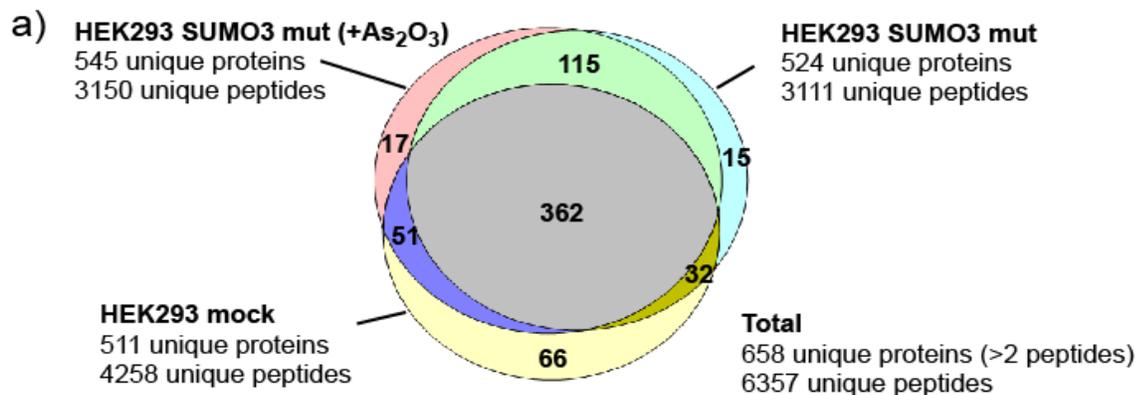


Figure 2.5: Mass spectrometry analyses of SUMOylated proteins from extracts of control HEK293 and HEK293 His<sub>6</sub>-SUMO-3 in As<sub>2</sub>O<sub>3</sub>-stimulated and non-stimulated cells. a) Venn diagram showing the overlapping distribution of proteins in each cell extract. Proteins identified with at least 2 peptides were considered for the comparison. b) Intensity distribution of peptide ions identified in the HEK293 His<sub>6</sub>-SUMO-3 in As<sub>2</sub>O<sub>3</sub>-stimulated and non-stimulated cells. The change in abundance of the PML tryptic peptide comprising the K490 residue is indicated on the scatter plot. c) MS/MS spectrum of precursor ion m/z 523.52<sup>4+</sup> using ETD fragmentation. Residues and c-type fragment ions characteristic of the SUMO-3 isopeptide bond are shown in red.

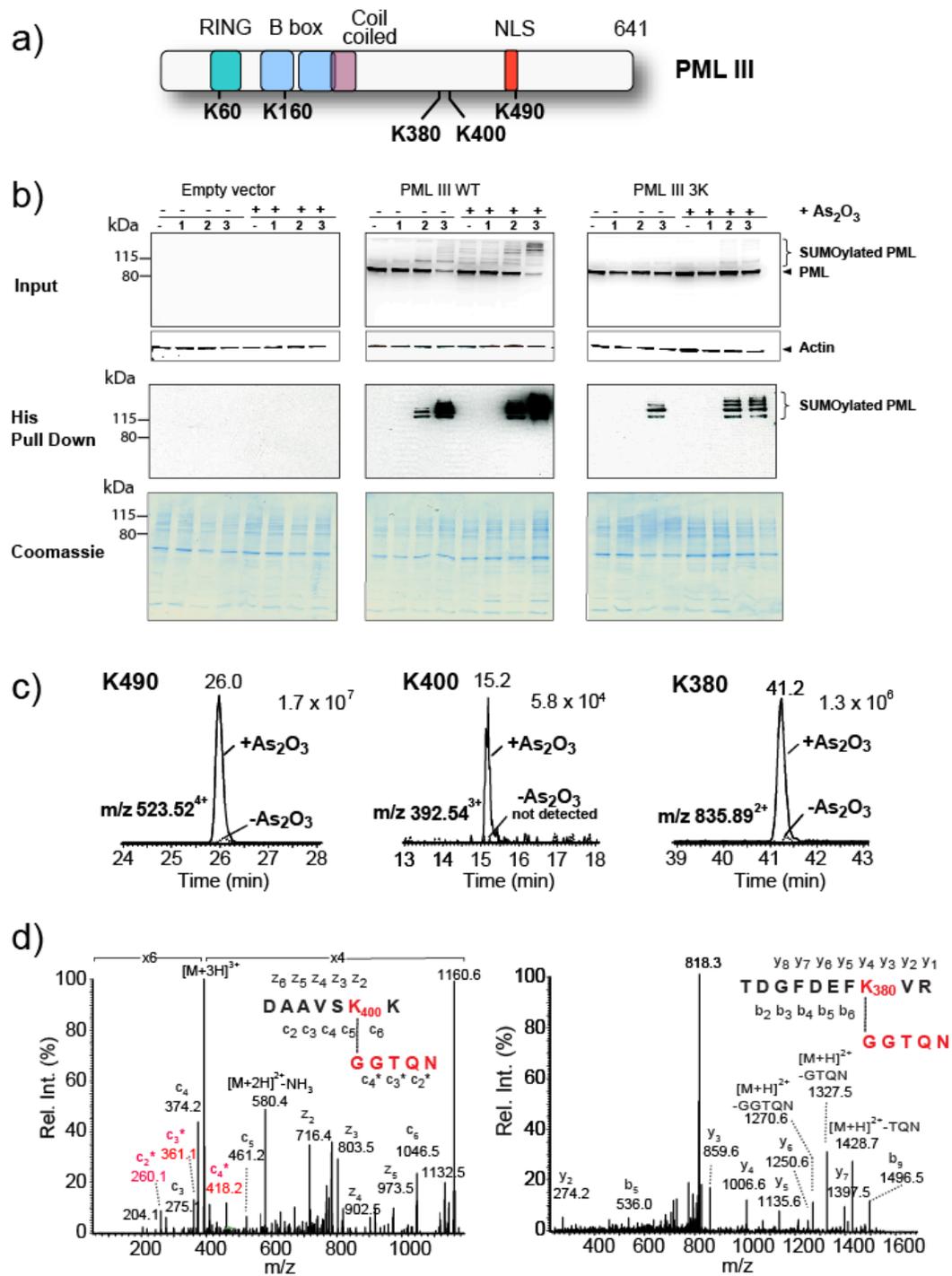
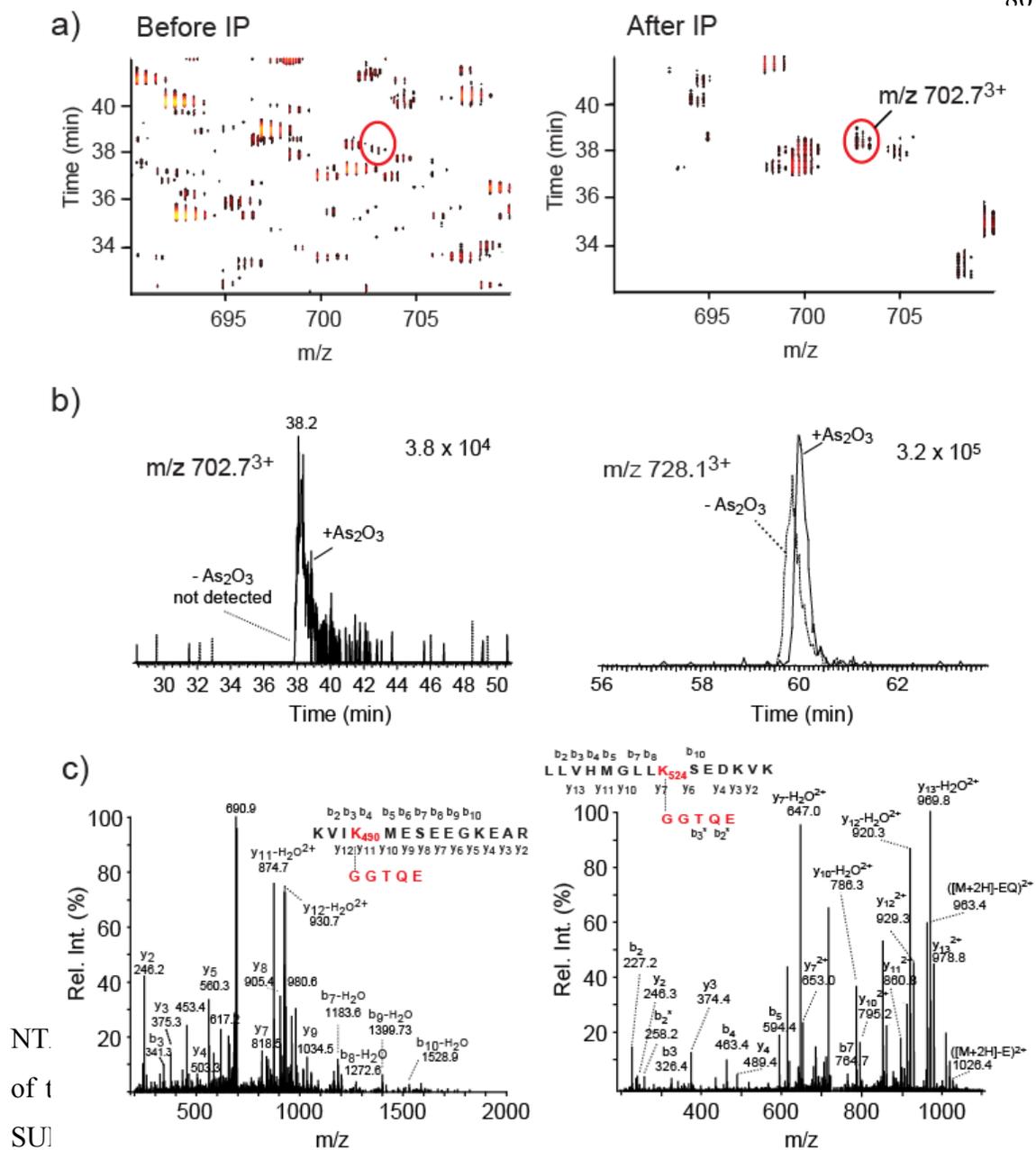


Figure 2.6: Identification of new SUMOylation sites in the protein promyelocytic leukemia (PML). a) Structure of PML III showing the protein domains and the location of known and novel SUMOylation sites. b) Immunoblots of protein extracts from HEK293 WT and HEK293-His-SUMO cells before (input) and after NTA enrichment (His pull down) revealed using an anti-PML antibody. The coomassie-stained SDS-PAGE gel is shown to compare sample loading. c) Extracted ion chromatograms for  $m/z$  523.52<sup>4+</sup>, 392.54<sup>3+</sup> and  $m/z$  835.89<sup>2+</sup> corresponding to modified K490, K401 and K380 PML tryptic peptides, respectively. d) MS/MS spectra of precursor ions  $m/z$  392.54<sup>3+</sup> and  $m/z$  835.89<sup>2+</sup> obtained using ETD and CID activation, respectively.



directed against the (K-GGTQE) epitope. b) Extracted ion chromatograms for  $m/z$  702.73+ and  $m/z$  903.83+ corresponding to modified tryptic peptides K490 from PML and K380 from SAFB2, respectively. c) MS/MS spectra of precursor ions  $m/z$   $m/z$  702.73+ and  $m/z$  903.83+ obtained using CID activation.

## 2.7 References

- Bernardi, R. and P. P. Pandolfi (2007). "*Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies.*" Nat Rev Mol Cell Biol **8**(12): 1006-16.
- Bernier-Villamor, V., D. A. Sampson, et al. (2002). "*Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1.*" Cell **108**(3): 345-56.
- Blomster, H. A., S. Y. Imanishi, et al. (2010). "*In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification.*" Journal of Biological Chemistry: -.
- Bossis, G. and F. Melchior (2006). "*Regulation of SUMOylation by Reversible Oxidation of SUMO Conjugating Enzymes.*" **21**(3): 349-357.
- Cooper, H. J., M. H. Tatham, et al. (2005). "*Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for the Analysis of Small Ubiquitin-like Modifier (SUMO) Modification: Identification of Lysines in RanBP2 and SUMO Targeted for Modification during the E3 AutoSUMOylation Reaction.*" Analytical Chemistry **77**(19): 6310-6319.
- Denuc, A. and G. Marfany "*SUMO and ubiquitin paths converge.*" Biochem Soc Trans **38**(Pt 1): 34-9.
- Desterro, J. M., M. S. Rodriguez, et al. (1998). "*SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation.*" Mol Cell **2**(2): 233-9.

- Everett, R. D. and M. K. Chelbi-Alix (2007). "*PML and PML nuclear bodies: implications in antiviral defence.*" Biochimie **89**(6-7): 819-30.
- Geiss-Friedlander, R. and F. Melchior (2007). "*Concepts in sumoylation: a decade on.*" Nat Rev Mol Cell Biol **8**(12): 947-956.
- Golebiowski, F., I. Matic, et al. (2009). "*System-wide changes to SUMO modifications in response to heat shock.*" Sci Signal **2**(72): ra24.
- Guo, B., S. H. Yang, et al. (2007). "*Signalling pathways and the regulation of SUMO modification.*" Biochem Soc Trans **35**(Pt 6): 1414-8.
- Guo, D., M. Li, et al. (2004). "*A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes.*" Nat Genet **36**(8): 837-41.
- Hay, R. T. (2005). "*SUMO: a history of modification.*" Mol Cell **18**(1): 1-12.
- Hayakawa, F. and M. L. Privalsky (2004). "*Phosphorylation of PML by mitogen-activated protein kinases plays a key role in arsenic trioxide-mediated apoptosis.*" Cancer Cell **5**(4): 389-401.
- Hietakangas, V., J. Anckar, et al. (2006). "*PDSM, a motif for phosphorylation-dependent SUMO modification.*" Proceedings of the National Academy of Sciences of the United States of America **103**(1): 45-50.
- Hoegel, C., B. Pfander, et al. (2002). "*RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO.*" Nature **419**(6903): 135-41.
- Hoeller, D. and I. Dikic (2009). "*Targeting the ubiquitin system in cancer therapy.*" Nature **458**(7237): 438-44.

- Hsiao, H.-H., E. Meulmeester, et al. (2009). "*ChopNspice*", a mass-spectrometric approach that allows identification of endogenous SUMO-conjugated peptides." Mol Cell Proteomics: M900087-MCP200.
- Kamitani, T., K. Kito, et al. (1998). "Identification of three major sentrinization sites in PML." J Biol Chem **273**(41): 26675-82.
- Kerscher, O., R. Felberbaum, et al. (2006). "Modification of proteins by ubiquitin and ubiquitin-like proteins." Annu Rev Cell Dev Biol **22**: 159-80.
- Lallemand-Breitenbach, V., M. Jeanne, et al. (2008). "Arsenic degrades PML or PML-RAR[alpha] through a SUMO-triggered RNF4/ubiquitin-mediated pathway." Nat Cell Biol **10**(5): 547-555.
- Lallemand-Breitenbach, V., J. Zhu, et al. (2001). "Role of promyelocytic leukemia (PML) sumolation in nuclear body formation, 11S proteasome recruitment, and As2O3-induced PML or PML/retinoic acid receptor alpha degradation." J Exp Med **193**(12): 1361-71.
- Lam, Y. W., A. I. Lamond, et al. (2007). "Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins." Curr Biol **17**(9): 749-60.
- Lin, D., M. H. Tatham, et al. (2002). "Identification of a substrate recognition site on Ubc9." J Biol Chem **277**(24): 21740-8.
- Marcantonio, M., M. Trost, et al. (2008). "Combined enzymatic and data mining approaches for comprehensive phosphoproteome analyses: application to cell signaling events of interferon-gamma-stimulated macrophages." Mol Cell Proteomics **7**(4): 645-60.

- Martin, S., K. A. Wilkinson, et al. (2007). "*Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction.*" Nat Rev Neurosci **8**(12): 948-959.
- Masclé, X. H., D. Germain-Desprez, et al. (2007). "*Sumoylation of the transcriptional intermediary factor Ibeta (TIF1beta), the Co-repressor of the KRAB Multifinger proteins, is required for its transcriptional activity and is modulated by the KRAB domain.*" J Biol Chem **282**(14): 10190-202.
- Matafora, V., A. D'Amato, et al. (2009). "*Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition.*" Mol Cell Proteomics **8**(10): 2243-55.
- Matic, I., J. Schimmel, et al. (2010). "*Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif.*" Mol Cell **39**(4): 641-652.
- Matic, I., M. van Hagen, et al. (2008). "*In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy.*" Mol Cell Proteomics **7**(1): 132-44.
- Matunis, M. J., E. Coutavas, et al. (1996). "*A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex.*" J Cell Biol **135**(6 Pt 1): 1457-70.
- Mukhopadhyay, D. and M. Dasso (2007). "*Modification in reverse: the SUMO proteases.*" Trends in Biochemical Sciences **32**(6): 286-295.

- Muller, S., M. J. Matunis, et al. (1998). "*Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus.*" Embo J **17**(1): 61-70.
- Owerbach, D., E. M. McKay, et al. (2005). "*A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation.*" Biochem Biophys Res Commun **337**(2): 517-20.
- Pampin, M., Y. Simonin, et al. (2006). "*Cross talk between PML and p53 during poliovirus infection: implications for antiviral defense.*" J Virol **80**(17): 8582-92.
- Pedrioli, P. G. A., B. Raught, et al. (2006). "*Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software.*" Nat Meth **3**(7): 533-539.
- Percherancier, Y., D. Germain-Desprez, et al. (2009). "*Role of SUMO in RNF4-mediated promyelocytic leukemia protein (PML) degradation: sumoylation of PML and phospho-switch control of its SUMO binding domain dissected in living cells.*" J Biol Chem **284**(24): 16595-608.
- Pichler, A., A. Gast, et al. (2002). "*The nucleoporin RanBP2 has SUMO1 E3 ligase activity.*" Cell **108**(1): 109-20.
- Pichler, A., P. Knipscheer, et al. (2005). "*SUMO modification of the ubiquitin-conjugating enzyme E2-25K.*" Nat Struct Mol Biol **12**(3): 264-9.
- Rogers, R. S., C. M. Horvath, et al. (2003). "*SUMO modification of STAT1 and its role in PIAS-mediated inhibition of gene activation.*" J Biol Chem **278**(32): 30091-7.

- Saitoh, H. and J. Hinchev (2000). "*Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3.*" J Biol Chem **275**(9): 6252-8.
- Schimmel, J., K. M. Larsen, et al. (2008). "*The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle.*" Mol Cell Proteomics **7**(11): 2107-22.
- Seeler, J.-S. and A. Dejean (2003). "*Nuclear and unclear functions of SUMO.*" Nat Rev Mol Cell Biol **4**(9): 690-699.
- Shen, L., M. H. Tatham, et al. (2006). "*SUMO protease SENP1 induces isomerization of the scissile peptide bond.*" Nat Struct Mol Biol **13**(12): 1069-1077.
- Stefan, R. W., K. Kirstin, et al. (2008). "*Arsenic trioxide stimulates SUMO-2/3 modification leading to RNF4-dependent proteolytic targeting of PML.*" FEBS letters **582**(21): 3174-3178.
- Swaney, D. L., G. C. McAlister, et al. (2008). "*Decision tree-driven tandem mass spectrometry for shotgun proteomics.*" Nat Meth **5**(11): 959-964.
- Tatham, M. H., M. C. Geoffroy, et al. (2008). "*RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation.*" Nat Cell Biol **10**(5): 538-46.
- Tatham, M. H., E. Jaffray, et al. (2001). "*Polymeric Chains of SUMO-2 and SUMO-3 Are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9.*" Journal of Biological Chemistry **276**(38): 35368-35374.
- Tatham, M. H., M. S. Rodriguez, et al. (2009). "*Detection of protein SUMOylation in vivo.*" Nat Protoc **4**(9): 1363-71.

- Vertegaal, A. C., J. S. Andersen, et al. (2006). "*Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics.*" Mol Cell Proteomics **5**(12): 2298-310.
- Wohlschlegel, J. A., E. S. Johnson, et al. (2006). "*Improved identification of SUMO attachment sites using C-terminal SUMO mutants and tailored protease digestion strategies.*" J Proteome Res **5**(4): 761-70.
- Yang, S. H., A. Galanis, et al. (2006). "*An extended consensus motif enhances the specificity of substrate modification by SUMO.*" Embo J **25**(21): 5083-93.
- Zhu, J., M. H. M. Koken, et al. (1997). "*Arsenic-induced PML targeting onto nuclear bodies: Implications for the treatment of acute promyelocytic leukemia.*" Proceedings of the National Academy of Sciences of the United States of America **94**(8): 3978-3983.

### **3. Discussion**

### 3.1 General Approach

SUMO is a PTM that attaches proteins via an isopeptide linkage through its C-terminal glycine and the  $\epsilon$  amino group of the target lysine. This modification is regulated by a specific set of conjugating and deconjugating enzymes (Ulrich 2008).

SUMOylation has been implicated in numerous important processes and diseases (Geiss-Friedlander and Melchior 2007). The main goal of this study is to develop a double affinity approach to selectively isolate SUMO-modified peptides bearing the modification site; to identify SUMO-conjugates peptides *in vivo* by mass spectrometry; and finally, to understand the effect of arsenic trioxide on this modification and, more precisely, on PML.

Through the use of HEK293 control cells or the stable expression of His-SUMO1/3 mutants, His-SUMO conjugates were purified in denaturing conditions with Ni-NTA beads. This approach proved to be successful in earlier reports for the study of Ub/Ubl conjugates (Peng, Schwartz et al. 2003). The mutants that were engineered can be discriminated by MS, rendering data interpretation easier compared to their wild type counterparts due to their shorter tag. Finally, the specific subset of SUMO-conjugated peptides bearing the SUMO-stub following trypsin digestion can be selectively immunoenriched.

### 3.2 *In Vitro* SUMOylation

The *in vitro* SUMOylation of proteins has become an indispensable tool for studying the functionality of this modification. The main advantage of *in vitro* systems is that it greatly simplifies data interpretation: the presence of a single substrate, the absence of competing reactions on lysines such as ubiquitylation or acetylation, and the absence of

SENPs. The recombinant conjugating enzymes (SAE1/SAE2, UBC9) and His-SUMO mutant or wild type are mixed with the substrate of choice in presence of Mg-ATP. SUMOylation results by a shift of ~20kDa; the outcome can therefore be separated by SDS-PAGE and viewed by silver staining or western blot against the substrate or the desired SUMO isoform. A control reaction is performed in the absence of Mg-ATP.

The *in vitro* SUMOylation reactions were conducted on numerous targets. Substrates bearing (RanGap) or not (E2-25K) the consensus motif were SUMOylated, and the results showed that all His-SUMO mutant isoforms exhibited a similar pattern to their wild type counterpart. Neither the His tag nor the mutation at the C-terminus of SUMO seems to interfere with the efficient recognition of conjugating enzymes, and the SUMOylation reaction is not impaired. The SUMOylation sites were also confirmed by mass spectrometry, and the peptides bearing the modification were detected, sequenced, and corresponded to the sites reported previously in the literature (Pichler, Knipscheer et al. 2005; Blomster, Imanishi et al. 2010).

SUMO-2/3 isoforms are known to form polySUMOylation chains *in vivo and in vitro*; however, the capacity of SUMO-1 to form polySUMOylation chains *in vivo* is still under investigation. Following the silver staining of the *in vitro* SUMO-1 reactions with E2-25K and RanGap in presence of ATP, it was possible to observe a banding pattern. We hypothesized that this pattern is probably due to polySUMO-1 chains, and this hypothesis was confirmed by mass spectrometry. Four polySUMOylation sites were detected and sequenced for SUMO-1: K23, K37, K39 and K48, two of which were previously observed by Cooper et al. (Cooper, Tatham et al. 2005), but none matched those identified by Pedrioli et al. (Pedrioli, Raught et al. 2006). It was intriguing to observe this discrepancy between different reports for the mapping of the SUMO1 polyconjugation sites *in vitro*. We concluded that this variability might be in part due to substrates used in different studies (different substrates might direct different poly-SUMO1 linkages). It is important to note that *in vitro* systems are artificial systems that are far from matching physiological

conditions, and that solvent exposed lysines could be artificially SUMOylated because of the high concentrations of conjugating enzymes. We concluded that although *in vitro* results can be a good tool in providing insight into the possible conjugation sites for the substrate of interest, one should always confirm the identified sites *in vivo*. However, we cannot rule out that this incongruity could also be due to the His-SUMO-1 mutant isoform we are using, which may have a slightly different 3D structure compared to its WT counterpart, thus rendering new lysines available for conjugation. Nonetheless, the consensus K11 poly-SUMO2/3 site that was observed *in vitro* when our mutant isoform was used agreed well with previous reports (Tatham, Jaffray et al. 2001).

The modifier being modified adds a new layer of complexity, but also versatility in biological function. It has been reported that different polyubiquitin linkages lead to different physiological functions (Pickart and Fushman 2004). Supposing the same applies to SUMO not only will we need to identify the specific SUMOylation site on the target, but also the type of polySUMO chain formed. Since SUMOylation has multiple functions, it is possible that the versatility of SUMO modification is due to the heterogeneity of the polySUMO chain formed. The heterogeneity does not only come from the lysine that is used to form linkages, but also from the modifier. As previously reported, it is possible to observe chains where ubiquitin is connected to other ubiquitin-like proteins such as NEDD or SUMO (Ikeda and Dikic 2008).

The need for a straightforward method in identifying SUMOylation sites *in vivo* became even more apparent due to the possible inconsistency that the *in vitro* conjugating system can introduce.

Two opposing processes regulate SUMOylation in the cell: the conjugation and deconjugation reactions. In order to further assess the functionality of the mutant SUMO, *in vitro* deSUMOylation reactions were conducted using His-SUMO1 wild type and mutant. First, *in vitro* SUMOylation reactions were conducted in presence and absence of Mg-ATP.

The E1 activity requires  $Mg^{2+}$ , the SUMOylation reaction is stopped with the addition of EDTA, a chelating agent that binds to  $Mg^{2+}$ . Following SENP1 treatment, the band corresponding to E2-25K-SUMO and RanGap-SUMO is absent for both the wild type and the mutant SUMO1 (Annex I, Supplementary Figure 2.3). The mutation and the His tag on SUMO do not interfere with SENP1 recognition and cleavage.

### 3.3 The His-SUMO Mutant Functionality in HEK293 Cells

The sequence of SUMO protein has no tryptic cleavage site in the vicinity of the C-terminus. Consequently, following trypsin digestion of a SUMOylated protein, 19 and 32 amino acids tags are released respectively for the human SUMO-1 and SUMO2/3. These large peptides are low abundance, and in addition their MS/MS interpretation is highly complex. To circumvent those challenges, His-SUMO mutants were created with a trypsin cleavage site at the 6<sup>th</sup> position from the C-terminus. Following trypsin digestion, a characteristic 5 amino acid stub is released.

*In vivo* functional assays by western blot and immunofluorescence were performed by F. Galisson to confirm the functionality of the mutant SUMOs. SUMO is known to localize in NBs or PML bodies (Kamitani, Kito et al. 1998). One of the most studied targets of SUMOylation is PML, mainly due to its role in APL (Lallemand-Breitenbach and de Thé 2010)., Treatment with  $As_2O_3$  triggers PML SUMOylation, which is subsequently ubiquitylated and then degraded (Zhang, Yan et al. 2010).

The functionality of SUMO mutants was tested on PML in the presence and absence of  $As_2O_3$ . Both the wild type and mutant SUMO show a similar pattern that consists in an increase in the MW corresponding to the mono- or polySUMO conjugated PML (Figure 2.3). We observe the depletion of PML following  $As_2O_3$  due to its subsequent ubiquitylation, which is consistent with previous reports (Zhu, Chen et al. 2002).

In the immunofluorescence experiment, briefly the goal was to visualize the effect of  $As_2O_3$  on the distribution of PML and His-SUMO wild type and mutant isoforms. Both the mutant and the wild type SUMO colocalize with the YFP-PML into nuclear bodies as previously reported (Lallemand-Breitenbach and de The 2010). This further confirms the functionality of the mutants *in vivo*.

### **3.4 Overexpression of His-SUMO Proteins in HEK293 Cells**

In order to facilitate SUMOylation site identification, the His-SUMO protein is overexpressed in HEK293 cells. The goal behind overexpression is to increase the level of endogenously low SUMO conjugates (Geiss-Friedlander and Melchior 2007). However, overexpressing SUMO can possibly lead to artifacts such as the SUMOylation of non-SUMOylable lysines. Moreover, overexpressing proteins can also lead to aggregate formation and their subsequent degradation (Sullivan 2007 ). It has also been reported that overexpression might potentially lead to misfolding; but, SUMO is a small protein (11kDa) and immunofluorescence studies have shown its proper localization in NBs. As described previously, the immunofluorescence studies that were performed by F. Galisson on the transfected cells showed a pattern of SUMO localization that reflected the reported pattern observed in literature (Kamitani, Kito et al. 1998).

There are multiple advantages in using HEK293 cells, notably the ease of transfection and growth for large-scale experiments. HEK293 are easily transfected through the use of calcium phosphate, an easy and inexpensive method. In large-scale experiments, up to 500 million cells per condition are lysed, and HEK293 cells have the advantage of growing fast and easily. Despite this fact, HEK293 cells are highly transformed and there is controversy over which cell type they are (Shaw, Morse et al. 2002).

In the present study, polyhistidine-tagged SUMOs are used: this enables the subsequent purification using nickel affinity columns. Previously, it has been shown that increasing the

number of tags on SUMO decreases the efficiency of conjugation (Hannich, Lewis et al. 2005). In this study, it is assumed that the histidine tag will not alter the protein property. The polyhistidine tag did not interfere with the *in vitro* SUMOylation or deSUMOylation assays, indicating that the recognition by E1, E2 and the SENP is not impaired. Moreover, it has been shown that polyhistidine tags show little effect on protein structure (Carson, Johnson et al. 2007). Based on these considerations, this study was conducted on the assumption that the polyhistidine tag had little to no effect on SUMOylation. Nonetheless, the recognition by E3 enzymes has not been tested, and if one considers working with a specific E3 ligase using the His-SUMO construct, proper *in vitro* SUMOylation control assays should be carried out.

In this study, we focused on SUMO-1 and SUMO-3 substrate identification. SUMO-2 was excluded for two main reasons. Firstly, SUMO-2 and SUMO-3 are known to be highly homologous (95%), and therefore it is assumed that they have similar targets. Secondly, the CID fragmentation of SUMO-2 conjugated peptides leads to a highly abundant ion corresponding to a loss of water (-18). This loss is strongly privileged in CID and thus compromises the fragmentation of the peptide. As illustrated in Figure 2.2b, the SUMO-2 mutant fragments are ~5 times lower in intensity by comparison to SUMO-1 and SUMO-3. This renders the MS/MS interpretation and sequencing more challenging.

Moreover, the transfected HEK293 cells co-express the His-SUMO mutant protein but also the endogenous SUMO. The parallel expression of both genes implies there is a competition between the two forms of SUMO for conjugation. To maximize the identification of SUMO conjugates, the sole expression of the His-SUMO mutant is preferred. Homologous recombination of the endogenous gene with the mutant SUMO is technically challenging.

In summary, site identification has been difficult since SUMOylation is of very low abundance (Pedrioli, Raught et al. 2006). Overexpressing SUMO in HEK293 cells is a

good starting point for the study of SUMOylation, even though the conditions are not physiological.

### 3.5 Subcellular Distribution of SUMO

SUMO is mainly nuclear, although in recent years cytoplasmic targets have been identified (Geiss-Friedlander and Melchior 2007). To enrich the sample in SUMO conjugated proteins, cells were fractionated prior to lysis, and the cytoplasmic and nuclear fractions were analyzed by western blot. As observed in Figure 2.4a, SUMO conjugates are highly enriched in the nucleus compared to the cytoplasm for SUMO-1 and SUMO-3. However, free or unconjugated SUMO is more abundant in the cytoplasm. Suggesting that upon conjugation SUMOylated proteins are shuffled to the nucleus; consistent with previous reports showing SUMOylation to be important for nuclear transport (Rodriguez, Dargemont et al. 2001; Munirathinam and Kalyanasundaram 2008).

The western blot signal of SUMO-2/3 is more important than that of SUMO-1. Two reasons might explain the higher signal for SUMO-2/3 by western blot: firstly, the higher level of SUMO-2/3 conjugates in the cell; secondly, that SUMO-2/3 is known to form polymeric chains *in vivo*, which is still under investigation for SUMO-1 (Tatham, Jaffray et al. 2001). However, we cannot rule out the possibility that the difference in western blot signal is mainly due to the formation of SUMO-2/3 polymeric chains, since western blot cannot differentiate between SUMO-conjugates or polySUMO chains. The mass spectrometry analysis identified no SUMO-1 conjugated peptides following the Ni-NTA (results not shown); while 19 SUMO-3 modified peptides were identified through the same approach. In conclusion, in HEK293 cells the SUMO-2/3 conjugates are in higher abundance compared to SUMO-1 conjugates as observed by western blot and mass spectrometry.

A successful enrichment method should i) enrich in SUMO targets, and ii) deplete in contaminant proteins. The first advantage of analyzing the nuclear fraction is the enrichment of SUMO conjugates as discussed earlier. The second is that the sample complexity is highly reduced. Indeed, about 25% of proteins in the cell are nuclear (Finn, Mistry et al. 2006); therefore if only the nuclear fraction is analyzed, the possible contamination by highly abundant cytoplasmic proteins is reduced. In fact, the presence of highly abundant proteins compromises the detection and the sequencing by mass spectrometry of low abundance proteins such as SUMO conjugates. In addition, the high abundance of the unconjugated His-SUMO in the cytoplasm could also compromise the nickel pull-down. Since free His-SUMO binds to the Ni-NTA column, free His-SUMO could compete with SUMOylated conjugates during the purification.

The main disadvantage of limiting to nuclear proteins is that we do not gain a global understanding of SUMOylation. Therefore future studies should also be carried out on cytoplasmic fraction.

### **3.6 Ni-NTA Enrichment**

The selective isolation of SUMO conjugates in HEK293 cells was performed using cells overexpressing His-SUMO1/3 and a control non-transfected HEK293 cell line. The nuclear extracts were lysed in denaturing conditions (6M guanidine) and the proteins were purified by affinity chromatography using Ni-NTA resin. The highly denaturing conditions minimize the copurification of proteins that are associated to SUMO conjugates. Furthermore, the denaturing conditions inactivate the SENPs present in the lysate. Other tag approaches could be used, such as HA, Myc or Flag tag. However, those approaches rely on antibody-antigen recognition, consequently, denaturing conditions cannot be used.

Following metal affinity chromatography of nontransfected cells, a high signal representing the background is observed (Annex I, Supplementary Figure 2.6). About the

same number of proteins (~400) were identified in the mock cells as in the His-SUMO transfected cell line. The contaminant proteins identified in the mock cells represent the background that is subtracted from the transfected cell line identifications. By analyzing the nature of the mock proteins, we identified three types of contaminants: i) highly abundant nuclear proteins such as the hnRNP; ii) metal-binding proteins such as zinc fingers that can also bind to nickel; and finally iii) proteins containing multiple histidines in their primary sequence such as Hox-A9. A major challenge in this type of study is to discriminate between background proteins and *bona fide* SUMO conjugates. In order to get a better insight into the nature of the background proteins following nickel pulldown, we suggest the creation of a repertoire of contaminant proteins as it was previously proposed for Tap tagging (Gingras, Aebersold et al. 2005).

### 3.7 Digestion and Peptide Separation

The vast majority of MS-based proteomics studies utilize trypsin to cleave proteins into more analyzable peptides (Olsen, Ong et al. 2004). Trypsin has the advantage of being highly specific to the C-terminus of arginines and lysines, active under a wide range of conditions, and of producing doubly charged peptides that are easily analyzed by CID (Olsen, Ong et al. 2004). The SUMO mutant was engineered to produce a 5 residue stub (EQTGG for SUMO-1) cross-linked to lysine following trypsin digestion. When the wild type SUMO-1 is digested a 19 amino acid stub is left on the acceptor lysine. However, it is also possible to obtain a small stub (QTGG) by digesting the wild type SUMO-1 with Glu-C. Despite the small SUMO tag left following the digestion, Glu-C is much less efficient in producing MS-analyzable peptides, in fact Swaney et al. showed a 36% decrease in peptide identification when Glu-C is used in comparison to trypsin on total yeast extract (Swaney, Wenger et al. 2010).

Prior to MS/MS analysis and following tryptic digestion, the peptides are separated by MudPIT (Multidimensional Protein Identification Technology), a two-dimensional liquid chromatography technique. First, the peptides are separated by charge with the use of a strong cation exchange phase, followed by reverse phase separation on a C18 column. Cation exchange is used because peptides are positively charged following trypsin digestion. C18 columns are universally used for peptide separation in large scale experiments because they have a range of interaction with a broad variety of compounds. Following the SUMO-3 study, ~6000 unique peptides were identified through this technique. Out of the 19 SUMO peptides identified, 8 contained a trypsin miscleavage adjacent to the SUMO-modified lysine. When there is a lysine or an arginine in the vicinity of the modified lysine, the SUMO modification seems to hinder trypsin from cleaving the peptide. Following this observation, we concluded that SUMO peptides are usually large due to the extra stub and multiple miscleavages, and thus hydrophobic. It is possible that some SUMO peptides were too hydrophobic to be analyzed by C18 chromatography. For instance, the SUMO peptide of PML K160 or K65 was not observed. Although the K160 site is known to be SUMOylated and regulated by arsenic (Lallemant-Breitenbach, Jeanne et al. 2008), the corresponding tryptic peptide has not been detected due to its hydrophobicity. Using the sequence-specific retention calculator tool, we calculated the theoretical retention time of the SUMOylated K160 peptide to be ~53 minutes (Krokhin 2006). The gradient stops at 55 minutes, therefore it is highly likely that the peptide never eluted. The identification of these hydrophobic peptides could possibly be improved by the use of a C5 column.

SUMO-modified peptides are branched with two N-termini. This unique aspect could be exploited to selectively isolate this class of peptides using FAIMS (High-Field Asymmetric Waveform Ion Mobility Spectrometry), which separates ions according to their ion mobility. If branched peptides have characteristic ion mobility, they could be isolated from the rest of the mixture.

## **3.8 SUMO Peptide Identification by Mass Spectrometry**

### **3.8.1 Sample Complexity**

Out of 6282 peptides identified in the SUMO-3 MudPIT experiment, 19 were confirmed to be SUMO peptides. The very proportion of identified SUMO peptides is explained by two reasons. Firstly, SUMO peptides are of low abundance in the sample, and although they can be detected in the MS spectrum, they are not selected for MS/MS due to the presence of high-abundance co-eluting species. The data-dependent acquisition method privileges the sequencing of high-abundance species. Since the ion elutes over a ~30 second time window, the instrument cannot acquire MS/MS spectra for all ions detected and as a result, low-abundance ions are not sequenced. This is exemplified in the large scale SUMO-1 study. In Figure 2.7a the SUMO-1 conjugated peptide of K490 PML is present before and after the IP at the same level. However, the peptide was only identified following the IP because the sample complexity is reduced. It is thus likely that in the large-scale SUMO-3 experiment, numerous SUMOylated peptides co-eluted with highly abundant species and consequently were not sequenced.

### **3.8.2 Spectral interpretation strategies**

The second main limitation in SUMO-peptide identification is the interpretation of the MS/MS spectrum. When the synthetic E2-25K K14 SUMO peptide is submitted to MASCOT, the search engine was not able to correctly sequence the peptide regardless the high quality of the MS/MS acquired (data not shown). Upon fragmentation SUMOylated peptide produces two ions series: one coming from the modification and one from the substrate peptide resulting in a mixed spectrum. It is expected that a certain proportion of SUMO modified peptide present in sample were not correctly assigned by MASCOT.

During a large scale experiment up to 100 000 MS/MS spectra are acquired, manual analysis of each MS/MS spectrum is impossible.

In order to get around the high error rate of MASCOT, signature fragments and neutral losses specific to SUMO stub (for instance, EQTGG for SUMO-1) were used to validate the *in vivo* spectra identified by MASCOT. Each mutant isoform of SUMO leads to a characteristic fragmentation; in the case of mutant SUMO-1 fragments at  $m/z$  258.1, 240.1 and 341.1 are specific to the EQTGG sequence in CID. Theoretically, all SUMO-1 peptides upon CID fragmentation should have these signature fragments. The same phenomenon is observed with ETD. However, since the fragmentation is performed in the linear trap the fragments that are smaller than the 1/3 of the precursor's  $m/z$  are not detected. In case where the precursor  $m/z$  was large ( $m/z > 3$  times the  $m/z$  of signature fragment), the characteristic fragment ions are not detected. We expect this phenomenon to be important since SUMO peptides are large peptide.

As observed by Jeram *et al* the fragmentation of the wild type SUMO modified peptide mainly leads to fragments coming from the SUMO chain and very little from the substrate peptide (Jeram, Srikumar et al. 2010). This phenomenon facilitates the identification of the modification, but unfortunately not of the substrate peptide and cannot reliably confirm the SUMOylation site. In contrast, the identification of mutant SUMO is facilitated. Because the mutant's chain is smaller, the number of SUMO specific fragments is significantly reduced permitting the use of standard sequencing software. As a result, MASCOT was the starting point in the interpretation of the MS/MS spectra. All SUMO modified peptides identified by MASCOT with a score  $>5$  were analyzed manually. Those having a relatively good quality MS/MS spectrum and the presence of signature ions were considered as true positives. To minimize the false discovery rate the score cut off is set to 25 in standard proteomic studies. However because of MASCOT's limitation in interpreting SUMO peptide spectrum, the score cut off is reduced. For instance, the H3 K24 SUMOylation site discovered in this study has a MASCO score of 13. Careful manual analysis of the

spectrum confirmed the identification. In some instances manual validation can also detect false positives. The SUMO peptide of RBM33 had a score of 48 with MASCOT, however after manual inspection, no SUMO specific neutral losses or signature ions were observed. Moreover, RBM33 SUMO modified peptide is also present in the negative control. In conclusion, manual validation of the MS/MS spectrum is a key step in the identification process.

ChopNSpice is a database search tool that allows for the search of branched peptide by linearizing the peptide *in silico* (Hsiao, Meulmeester et al. 2009). Using this approach we identified one important site not observed using MASCOT: the K13 of H4 (Annex IV, Supplementary Table III). ChopNSpice database is 10X larger than the standard database, increasing the analysis time. Moreover, the linearization of the peptide introduces an artifact: in fact when the modification is located close to the C-terminus of the peptide, ChopNSpice performs poorly. For instance PML K497 (DAAVSK(SUMO)K) peptide was uniquely identified by MASCOT.

Summon algorithm was also tested, but the identification rate was low (Annex IV, Supplementary Table III). Summon is optimal for large tags, such as the wild type SUMO. In contrast to the wild type SUMO, the mutant SUMO is shorter and therefore results in a fewer fragments. Summon limitation for shorter tags has been reported where it performed worst for the yeast isoform of SUMO, SMT3 (5 amino acid long tag) compared to human SUMO-1 (19 amino acid long tag) (Jeram, Srikumar et al. 2010).

In conclusion, automated identification of SUMOylation sites is not yet optimal and manual validation is required. Further developments in search algorithms should enhance the sensitivity of SUMOylation site discovery.

### 3.8.3 Fragmentation methods

Two different fragmentation techniques were used in the large scale SUMO-1 and SUMO-3 studies. CID and ETD were used in alternation in a decision tree manner based on the charge and the  $m/z$  of the precursor (Swaney, McAlister et al. 2008). CID is the traditional method mainly used for doubly charged species, and in recent years ETD has emerged as an important alternative fragmentation technique which is more optimal for highly charged species. It has been observed that SUMO modified peptide mainly exists in charge states  $> +3$  due to their additional N-terminus. Although the majority of SUMO-3 peptides were identified by both fragmentation modes or by CID alone, the K497 of PML and K420 of Lamin were uniquely identified using ETD (Annex IV, Supplementary Table III). By combining both modes the number of identification is increased as well as the confidence in those peptides that were sequenced by both modes. The main limitation of CID and ETD is the low mass resolution of fragment ions and the loss of low  $m/z$  fragments since which are not trapped in the LTQ. One way to obtain low  $m/z$  information and increase the mass resolution of product ion spectra is by HCD. Although HCD can bring multiple advantages, a high reduction in sensitivity has been reported (Dayon, Pasquarello et al. 2010).

### 3.9 Immunoprecipitation and SUMO-1 Site Identification

A monoclonal antibody was generated in mouse against a synthetic peptide containing a lysine residue modified with the SUMO-1 stub. 14 different antibodies were generated and the best selected based on its specificity and selectivity against synthetic SUMO peptides of PML, RanGap and E2-25K. An immunoprecipitation assay was developed using a centrifugal filter device. Upon application of centrifugal force SUMOylated peptides bound to the antibody are retained in the upper chamber while the non specific peptides are collected in the bottom chamber (Annex I, Supplementary Figure 2.8a). Since most tryptic

peptides are <3kDa, they pass through the 10kDa filter unless bound to the anti-EQTGG antibody.

The antibody was tested using synthetic peptides at varying concentrations in a tryptic digest of HEK293 cells at a fixed antibody concentration (Annex I, Supplementary Figure 2.8b). The general trend observed is that the recovery increases with a decrease in SUMO peptide. At 5 fmol the recovery of E2-25K is the highest (75%), however the PML K490 peptide was not detected at 5 fmol probably because it is below the limit of detection of this peptide. As the amount of antigen increases, while the antibody concentration stays constant, the ratio antibody:SUMOylated peptide decreases and as a consequence the recovery drops. Because of the limited amount of antibody, the optimal antibody: antigen ratio for large scale studies is 10:1. At this ratio, the recovery is around ~30%. Based on these results and taking into account the limit of detection of most synthetic SUMO peptides (around 20 fmol), the SUMOylated peptides should be at least at 60 fmol in the initial *in vivo* extract prior to the IP to be detected in the elution sample. The analysis of the SUMO-1 Ni-NTA extract did not identify any SUMO peptide. However, following the immunoprecipitation 3 SUMOylated peptides of RanGap, PML and SAFB2 were detected. A large number of non SUMOylated peptides were present in the elution, in the SUMO-1 As<sub>2</sub>O<sub>3</sub> treated sample 408 peptides were identified by MASCOT and only 3 were SUMOylated. The low proportion of SUMOylated peptide in the elution is probably due to multiple factors. First, the specificity of the antibody is poor because it was directed against a small sequence (EQTGG). We observed a high proportion of glutamic acid (E) containing peptides compared to the natural abundance. In fact, glutamic acid was found at 12.4% in our elution sample compared to its natural abundance of 6.3% (Betts and Russell 2003). This contrasts to alanine which was found at an abundance of 7.4% in our sample compared to 7.8% in nature (Betts and Russell 2003). Therefore, the antibody seems to preferentially bind to peptides rich in glutamic acid. Our hypothesis is that this sequence is too short to get a very specific immunogenic reaction in mice. In fact, the optimal epitope length is approximately 15 amino acids (Benjamin and Perdue 1996). The issue of low

antibody specificity could be resolved by using a more specific antigen. Xu et al. developed a similar antibody approach but with a very high specificity against the di-glycine tag specific to ubiquitin (Xu, Paige et al. 2010). In contrast to our approach, the antigen used was a protein with multiple di-glycines. As a result, the antibody recognized di-glycine tagged proteins and peptides with a very high specificity. Therefore, a protein antigen with multiple lysines all modified with EQTGG could increase the specificity of the antibody.

A potent deSUMOylase inhibitor could increase the level of SUMO conjugates in the cell. The ubiquitin studies use MG132, a proteasome inhibitor, resulting in an accumulation of ubiquitin conjugates in the cell prior to the lysis. No SENP specific inhibitor was developed to permit a similar accumulation of SUMO conjugates. Recently it was shown that MG132 treatment also increases the level of SUMO conjugates in the cell (Bailey and O'Hare 2005), probably because SUMO can be recognized by ubiquitin ligases and serve as a proteolytic targeting signal (Uzunova, Gottsche et al. 2007). Although a good proportion of SUMO conjugates accumulate upon MG132 treatment, it has been shown that SAFB2 SUMOylation decreases upon proteasome inhibition (Schimmel, Larsen et al. 2008).

### **3.10 Effect of Arsenic Trioxide**

Western blot and mass spectrometry results showed an increase in SUMOylation of PML upon arsenic treatment as reported in literature (Stefan, Kirstin et al. 2008). Our MS method is reliable to identify and quantify SUMOylation *in vivo*. The advantage of mass spectrometry is the possibility to monitor the regulation of different site on the same protein (Figure 2.6c). A possible artifact of quantification is the presence of polySUMO chains. The polySUMOylated proteins are more efficiently enriched by Ni-NTA because of multiple poly histidine tags.

Arsenic trioxide is known to induce SUMOylation of PML and 3 of its conjugation sites are known (K65, 160 and 490) (Kamitani, Kito et al. 1998). PML-R $\alpha$ R $\alpha$  is a fusion protein expressed in APL which disrupts NBs, but As<sub>2</sub>O<sub>3</sub> treatment promotes PML-R $\alpha$ R $\alpha$  degradation and restores NBs (Zhu, Koken et al. 1997). It is through K160 SUMOylation that arsenic promotes PML degradation (Lallemand-Breitenbach, Zhu et al. 2001). The increase in SUMOylation of PML is more important for SUMO-3 > SUMO-2 > SUMO-1 by western blot. This pattern is explained by the fact that SUMO-3 and SUMO-2 form polySUMOylation chains *in vivo*. PML served as a positive control in our study since its regulation by arsenic is known. One objective was to determine whenever PML is the sole target of over-SUMOylation following arsenic. To do so, data was processed by ProteoProfile: peptide identifications and abundances were clustered across different samples. Scatter plot of intensities with and without arsenic were constructed (for SUMO-3 Figure 2.5b). In the SUMO-3 sample, the abundance of the vast majority of peptides identified (92%) was unaffected and the most significant change is for PML: 15 fold increase in SUMOylation in presence of As<sub>2</sub>O<sub>3</sub>. The K160 and K65 sites of PML were not identified in this study, probably due to their important size. The identification of large peptides is impaired due to chromatographical and MS/MS interpretation challenges. The K490 PML SUMOylation site was identified for SUMO-1 and 3 and is induced by arsenic. PML seems to be the main target of arsenic and therefore we can speculate that As<sub>2</sub>O<sub>3</sub> treatment in APL is specific to PML and PML-R $\alpha$ R $\alpha$ . The scatter plot of SUMO-1, revealed that as expected RanGap K524 SUMO-peptides is not induced by arsenic, while PML K490 was only detected in presence of arsenic. In addition, the western blot anti-PML is super impossible to western blot anti-SUMO following Ni-NTA enrichment, further suggesting that the main target of arsenic is PML (Figure 2.4b).

Studies performed on the PML 3K (where all SUMOylable lysines are mutated) showed residual SUMOylation (Figure 2.6b). The lower signal observed PML 3K SUMOylation suggests that the major modification sites are the known 3 sites. Nonetheless, these results suggested unknown sites in PML. The mass spectrometry

analysis of SUMO-3 extracts revealed the presence of 3 new sites (K380, 400 and 497) in PML all upregulated by arsenic (Figure 2.6c). The identity of PML K380 and K400 SUMOylation was confirmed by synthesizing the corresponding peptides. The tandem spectra of the synthetic peptides gave a highly similar fragmentation pattern to the *in vivo* spectra confirming the MASCOT identifications (Annex I, Supplementary Figure 2.7). The tandem spectrum of K497 was not shown due to its complexity: in fact the peptide is also SUMOylated at K490 rendering interpretation more complex. The new sites identified are known ubiquitylation sites (Tatham, Geoffroy et al. 2008), possibly competing with ubiquitylation. This mechanism contrasts with the SUMOylation of K160 where the SUMO is itself ubiquitinated by RNF4 (Tatham, Geoffroy et al. 2008).

In conclusion, the mass spectrometry experiment detected and identified known (PML K490) and new (PML K380, 400 and 497) SUMOylation sites in a quantitative fashion.

### 3.11 Identified targets

The GO term analysis of SUMO-3 and SUMO-1 protein identification sample revealed common biological processes (Annex I, Supplementary Figure 2.5): chromatin maintenance and establishment and nucleic acid metabolism consistent with previous studies (Matafora, D'Amato et al. 2009). SUMO is known to regulate transcription and genome stability (Eilebrecht, Smet-Nocca et al. 2010). Moreover, SUMO-1 has been shown to directly bind to DNA in a sequence independent fashion. For example conjugation to transcription regulators could possibly compete for the DNA binding (Eilebrecht, Smet-Nocca et al. 2010). SUMO modification may play a similar role when conjugated to SAFB2 hypothetically competing for DNA binding. Identified proteins found in the chromatin establishment and maintenance group are histones and two new SUMOylation sites *in vivo* for H3 and H4 were mapped. Core histones are known to be “master control switches” (Zheng and Hayes 2003). Histone sites identified are known acetylation sites suggesting a possible interplay between SUMOylation and acetylation.

The crosstalk of SUMOylation and acetylation on numerous targets was previously reported (Choudhary, Kumar et al. 2009; Wu and Chiang 2009). For instance, both SUMOylation and acetylation target the same lysines but have opposing roles for PLAG1 and PLAGL2 (Zheng and Yang 2005). H4 SUMOylation has been previously suggested to be involved in transcriptional repression (Shiio and Eisenman 2003), and our study mapped the precise SUMOylation site of H4.

The SAFB2 is proposed to be a tumor suppressor gene in breast cancer (Townson, Dobrzycka et al. 2003). SAFB2 binds to the Scaffold/matrix attachment regions of DNA, those regions are important in gene expression regulation. Two SUMOylation sites *in vivo* for SAFB2, interestingly the two sites differed: K294 for SUMO-1 and K524 for SUMO-3. This demonstrates the heterogeneity of SUMO1 versus SUMO2/3 targets.

Lamin A, a highly modified protein and implicated in numerous diseases, is SUMOylated at K203 and the mutation of the consensus site is directly linked to familial dilated cardiomyopathy (Zhang and Sarge 2008). In this study, we mapped a new SUMOylation site for Lamin A/C the lysine 420. Further biological studies should be performed to understand the function of K420 SUMOylation.

Finally, novel mixed polySUMO linkages were identified. We identified SUMO-4 modified by SUMO-3 at K11. SUMO-4 is mainly expressed in kidney (Bohren, Nadkarni et al. 2004); therefore it is not surprising that SUMO-4 peptide is observed in HEK293 cells which are kidney cells. SUMO-4 contains a proline residue in the proximity of the diglycine tag that impedes its recognition by SENPs (Ulrich 2008). As a result, SUMO-4 conjugation is still under debate, some studies suggesting its conjugation only under cellular stress (Wei, Yang et al. 2008). We found that SUMO-4 is cross-linked at K11, the same lysine where SUMO-2 and SUMO-3 were found to be modified. In conclusion, although SUMO-4 role as a modifier is debated, it can be itself modified by SUMO-3, and

it may have other biological role in the cell, probably through non-covalent interactions if it is not through covalent modification of targets.

New polySUMOylation sites were identified for SUMO-2 and SUMO-3 protein: K42 and K41 respectively. These new sites were previously observed *in vitro* (Jeram, Srikumar et al. 2010), but there was no *in vivo* evidence. As previously demonstrated for ubiquitin (Pickart and Fushman 2004), different cross links could presumably lead to different physiological functions. These new sites might explain the diverse array of functions SUMO plays.

### 3.12 References

- Bailey, D. and P. O'Hare (2005). "Comparison of the SUMO1 and ubiquitin conjugation pathways during the inhibition of proteasome activity with evidence of SUMO1 recycling." Biochem J **392**(Pt 2): 271-281.
- Benjamin, D. C. and S. S. Perdue (1996). "Site-Directed Mutagenesis in Epitope Mapping." Methods **9**(3): 508-515.
- Betts, M. J. and R. B. Russell (2003). Amino Acid Properties and Consequences of Substitutions, John Wiley & Sons, Ltd.
- Blomster, H. A., S. Y. Imanishi, et al. (2010). "In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification." J Biol Chem: -.
- Bohren, K. M., V. Nadkarni, et al. (2004). "A M55V polymorphism in a novel SUMO gene (SUMO-4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus." J Biol Chem **279**(26): 27233-27238.
- Carson, M., D. H. Johnson, et al. (2007). "His-tag impact on structure." Acta Crystallographica Section D **63**(3): 295-301.
- Choudhary, C., C. Kumar, et al. (2009). "Lysine acetylation targets protein complexes and co-regulates major cellular functions." Science **325**(5942): 834-840.
- Cooper, H. J., M. H. Tatham, et al. (2005). "Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for the Analysis of Small Ubiquitin-like Modifier (SUMO)

Modification: Identification of Lysines in RanBP2 and SUMO Targeted for Modification during the E3 AutoSUMOylation Reaction." Analytical Chemistry **77**(19): 6310-6319.

Dayon, L., C. Pasquarello, et al. (2010). "Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags." Journal of Proteomics **73**(4): 769-777.

Eilebrecht, S., C. Smet-Nocca, et al. (2010). "SUMO-1 possesses DNA binding activity." BMC Res Notes **3**(1): 146.

Finn, R. D., J. Mistry, et al. (2006). "Pfam: clans, web tools and services." Nucleic Acids Res(34 Database): D247 - 251.

Geiss-Friedlander, R. and F. Melchior (2007). "Concepts in sumoylation: a decade on." Nat Rev Mol Cell Biol **8**(12): 947-956.

Gingras, A. C., R. Aebersold, et al. (2005). "Advances in protein complex analysis using mass spectrometry." J Physiol **563**(Pt 1): 11-21.

Hannich, J. T., A. Lewis, et al. (2005). "Defining the SUMO-modified proteome by multiple approaches in *Saccharomyces cerevisiae*." J Biol Chem **280**(6): 4102-4110.

Hsiao, H.-H., E. Meulmeester, et al. (2009). ""ChopNspice", a mass-spectrometric approach that allows identification of endogenous SUMO-conjugated peptides." Mol Cell Proteomics: M900087-MCP900200.

- Ikeda, F. and I. Dikic (2008). "Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series." EMBO Rep **9**(6): 536-542.
- Jeram, S. M., T. Srikumar, et al. (2010). "An improved SUMmOn-based methodology for the identification of ubiquitin and ubiquitin-like protein conjugation sites identifies novel ubiquitin-like protein chain linkages." Proteomics **10**(2): 254-265.
- Kamitani, T., K. Kito, et al. (1998). "Characterization of a second member of the sentrin family of ubiquitin-like proteins." J Biol Chem **273**(18): 11349-11353.
- Kamitani, T., K. Kito, et al. (1998). "Identification of Three Major Sentrinization Sites in PML." J Biol Chem **273**(41): 26675-26682.
- Krokhin, O. V. (2006). "Sequence-Specific Retention Calculator. Algorithm for Peptide Retention Prediction in Ion-Pair RP-HPLC: Application to 300- and 100-Å Pore Size C18 Sorbents." Analytical Chemistry **78**(22): 7785-7795.
- Lallemand-Breitenbach, V. and H. de The (2010). "PML nuclear bodies." Cold Spring Harb Perspect Biol **2**(5): a000661.
- Lallemand-Breitenbach, V., M. Jeanne, et al. (2008). "Arsenic degrades PML or PML-RAR[alpha] through a SUMO-triggered RNF4/ubiquitin-mediated pathway." Nat Cell Biol **10**(5): 547-555.
- Lallemand-Breitenbach, V. r., J. Zhu, et al. (2001). "Role of Promyelocytic Leukemia (Pml) Sumolation in Nuclear Body Formation, 11s Proteasome Recruitment, and

as2O3-Induced Pml or Pml/Retinoic Acid Receptor  $\hat{I}\pm$  Degradation." The Journal of Experimental Medicine **193**(12): 1361-1372.

Matafora, V., A. D'Amato, et al. (2009). "Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition." Mol Cell Proteomics **8**(10): 2243-2255.

Munirathinam, G. and R. Kalyanasundaram (2008). "Sumoylation is important for the nuclear transport and antioxidant function of the translationally controlled tumor protein." FASEB J. **22**(1\_MeetingAbstracts): 1026.1011-.

Olsen, J. V., S.-E. Ong, et al. (2004). "Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues." Molecular & Cellular Proteomics **3**(6): 608-614.

Pedrioli, P. G. A., B. Raught, et al. (2006). "Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software." Nat Meth **3**(7): 533-539.

Peng, J., D. Schwartz, et al. (2003). "A proteomics approach to understanding protein ubiquitination." Nat Biotech **21**(8): 921-926.

Pichler, A., P. Knipscheer, et al. (2005). "SUMO modification of the ubiquitin-conjugating enzyme E2-25K." Nat Struct Mol Biol **12**(3): 264-269.

Pickart, C. M. and D. Fushman (2004). "Polyubiquitin chains: polymeric protein signals." Current Opinion in Chemical Biology **8**(6): 610-616.

- Rodriguez, M. S., C. Dargemont, et al. (2001). "SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting." J Biol Chem **276**(16): 12654-12659.
- Schimmel, J., K. M. Larsen, et al. (2008). "The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle." Mol Cell Proteomics **7**(11): 2107-2122.
- Shaw, G., S. Morse, et al. (2002). "Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells." FASEB J.: 01-0995fje.
- Shiio, Y. and R. N. Eisenman (2003). "Histone sumoylation is associated with transcriptional repression." Proc Natl Acad Sci U S A **100**(23): 13225-13230.
- Stefan, R. W., K. Kirstin, et al. (2008). "Arsenic trioxide stimulates SUMO-2/3 modification leading to RNF4-dependent proteolytic targeting of PML." FEBS letters **582**(21): 3174-3178.
- Sullivan, K. F. (2007 ). Fluorescent Proteins, Elsevier Science.
- Swaney, D. L., G. C. McAlister, et al. (2008). "Decision tree-driven tandem mass spectrometry for shotgun proteomics." Nat Meth **5**(11): 959-964.
- Swaney, D. L., C. D. Wenger, et al. (2010). "Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics." Journal of Proteome Research **9**(3): 1323-1329.

- Tatham, M. H., M. C. Geoffroy, et al. (2008). "RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation." Nat Cell Biol **10**(5): 538-546.
- Tatham, M. H., E. Jaffray, et al. (2001). "Polymeric Chains of SUMO-2 and SUMO-3 Are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9." J Biol Chem **276**(38): 35368-35374.
- Townson, S. M., K. M. Dobrzycka, et al. (2003). "SAFB2, a new scaffold attachment factor homolog and estrogen receptor corepressor." J Biol Chem **278**(22): 20059-20068.
- Ulrich, H. D. (2008). "The Fast-Growing Business of SUMO Chains." Molecular Cell **32**(3): 301-305.
- Ulrich, H. D. (2008). The SUMO System: An Overview. **497**: 3-16.
- Uzunova, K., K. Gottsche, et al. (2007). "Ubiquitin-dependent Proteolytic Control of SUMO Conjugates." J Biol Chem **282**(47): 34167-34175.
- Wei, W., P. Yang, et al. (2008). "A stress-dependent SUMO4 sumoylation of its substrate proteins." Biochemical and Biophysical Research Communications **375**(3): 454-459.
- Wu, S.-Y. and C.-M. Chiang (2009). "Crosstalk between sumoylation and acetylation regulates p53-dependent chromatin transcription and DNA binding." EMBO J **28**(9): 1246-1259.

- Xu, G., J. S. Paige, et al. (2010). "Global analysis of lysine ubiquitination by ubiquitin remnant immunoprecipitation." Nat Biotechnol **28**(8): 868-873.
- Zhang, X.-W., X.-J. Yan, et al. (2010). "Arsenic Trioxide Controls the Fate of the PML-RAR $\alpha$  Oncoprotein by Directly Binding PML." Science **328**(5975): 240-243.
- Zhang, Y. Q. and K. D. Sarge (2008). "Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies." J Cell Biol **182**(1): 35-39.
- Zheng, C. and J. J. Hayes (2003). "Structures and interactions of the core histone tail domains." Biopolymers **68**(4): 539-546.
- Zheng, G. and Y.-C. Yang (2005). "Sumoylation and Acetylation Play Opposite Roles in the Transactivation of PLAG1 and PLAGL2." J Biol Chem **280**(49): 40773-40781.
- Zhu, J., Z. Chen, et al. (2002). "How acute promyelocytic leukaemia revived arsenic." Nat Rev Cancer **2**(9): 705-713.
- Zhu, J., M. H. M. Koken, et al. (1997). "Arsenic-induced PML targeting onto nuclear bodies: Implications for the treatment of acute promyelocytic leukemia." Proceedings of the National Academy of Sciences of the United States of America **94**(8): 3978-3983.

## **4. Conclusion**

## 4.1 The SUMO challenge

Protein SUMOylation occurs on a wide variety of targets and has been implicated in numerous diseases (Geiss-Friedlander and Melchior 2007). The importance and versatility of SUMOylation in the cell is probably underestimated because of the difficulty in identifying SUMO substrates and more importantly their precise SUMOylation site.

The need for a reliable, unbiased and fast technique for identifying targets and their attachment site became apparent. We developed a unique *in vivo* MS based quantitative approach that permitted the identification of 17 and 3 SUMOylation sites for SUMO-3 and SUMO-1 respectively. Our approach relies on the enrichment of SUMOylated proteins in a similar manner as described for ubiquitin by Peng et al. (Peng, Schwartz et al. 2003), followed by the immunoprecipitation of peptides bearing the modification site. A double affinity approach is necessary for studying SUMOylation because of its very low abundance and high turn-over in the cell. The encouraging results obtained in this study as well as future ideas leads us to think that an even greater number of SUMOylation sites can be identified.

We identified known and novel SUMOylation sites for SUMO-3 and SUMO-1. The identification of known sites confirmed that our technique, although not exhaustive, is reliable. The identification of unknown SUMOylation sites, even for highly studied protein such as PML and histones, illustrates the unbiased nature of mass spectrometry. All SUMOylation sites identified on PML were induced by arsenic. As of our knowledge, we are the first to report an MS based quantitative tool to study site specific SUMOylation.

## 4.2 Future Perspectives

Our study focused on nuclear proteins, since a higher proportion of SUMO substrates is nuclear. Since recent reports highlighting the importance of SUMO outside the nucleus, we believe that the next large scale studies should also be performed on the cytoplasmic fraction. It will be interesting to evaluate the distribution of SUMO targets in different cell compartments. However, the analysis of cytoplasmic fraction will present further challenges due to the higher complexity of the sample and the lower abundance of the analyte. To overcome this challenge, further fractionation and enrichment of the sample will be required.

It is also important to increase the level of SUMO substrate in the cell; inhibiting the proteasome was shown to be successful (Matafora, D'Amato et al. 2009). Other agents are known to increase the level of SUMOylation such as oxidative stress and heat shocks.

Finally, although multiple E3 ligases were identified recently for SUMOylation, the pairing between the specific ligase and its substrates is challenging. As future work, it is possible to use the tools that we developed to identify specific E3 ligase targets. For instance, PIAS is an important E3 ligase known to promote the SUMOylation of multiple key proteins such as STAT and p53 (Palvimo 2007), but it is believed that an even greater number of PIAS specific targets can be identified (Rytinki, Kaikkonen et al. 2009). In fact, PIAS functions are believed to be due to its capacity to act as SUMO E3 ligase. E3 specific study at the proteome level could further reveal the function of PIAS in the cell.

We established the reliability of our method by clearly showing at multiple levels the agreement with previous reports. In addition, the discovery of unknown sites and its quantitative quality gives it multiple advantages as a leading tool in studying site specific SUMOylation.

### 4.3 References

- Geiss-Friedlander, R. and F. Melchior (2007). "Concepts in sumoylation: a decade on." Nat Rev Mol Cell Biol **8**(12): 947-956.
- Matafora, V., A. D'Amato, et al. (2009). "Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition." Mol Cell Proteomics **8**(10): 2243-2255.
- Palvimo, J. J. (2007). "PIAS proteins as regulators of small ubiquitin-related modifier (SUMO) modifications and transcription." Biochem Soc Trans **35**(Pt 6): 1405-1408.
- Peng, J., D. Schwartz, et al. (2003). "A proteomics approach to understanding protein ubiquitination." Nat Biotech **21**(8): 921-926.
- Rytinki, M., S. Kaikkonen, et al. (2009). "PIAS proteins: pleiotropic interactors associated with SUMO." Cellular and Molecular Life Sciences **66**(18): 3029-3041.

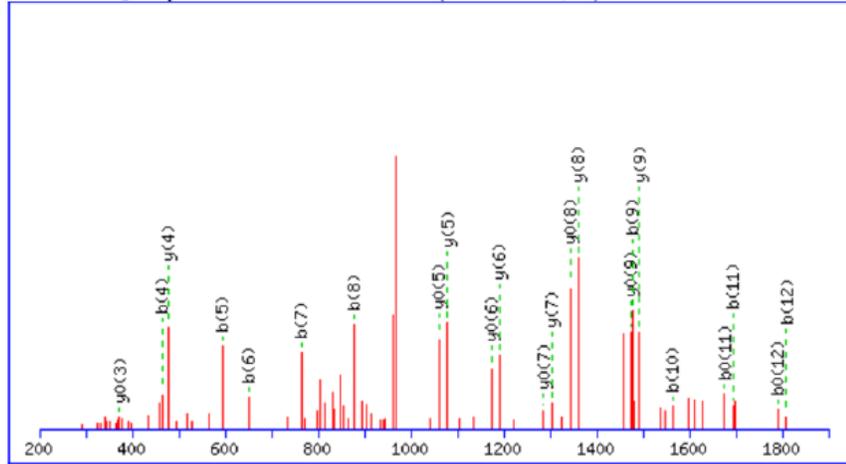
**Annex I: Supplementary Figures 2.1 -2.9 for Chapter 2**  
**(A Novel Proteomics Approach to Identify SUMOylated Proteins**  
**and their Modification Sites in Human Cells)**

a)

**MS/MS Fragmentation of LLVHMGLLKSEDK**

Found in In\_vitro\_RanGap\_Sumo1\_withATP\_12March\_2010\_bis\_100317173334.raw - IPI:IPI00294879-  
Tax\_Id=9606 Gene\_Symbol=RANGAP1 Ran GTPase-activating protein 1

Match to Query 1709: 1954.017428 from(978.015990,2+)



Monoisotopic mass of neutral peptide Mr(calc): 1954.02 Variable modifications: K9 : SUMO EQTGG  
(K) (K) Ions Score: 66 Expect: 0.0001 Matches (**Bold Red**): 23/108 fragment ions using 35 most intense peaks

#	b	b <sup>++</sup>	b*	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	114.09	57.55					L							13
2	227.18	114.09					L	1841.94	921.47	1824.92	912.96	1823.93	912.47	12
3	326.24	163.63					V	1728.86	864.93	1711.83	856.42	1710.85	855.93	11
4	<b>463.30</b>	232.16					H	1629.79	815.40	1612.76	806.89	1611.78	806.39	10
5	<b>594.34</b>	297.68					M	<b>1492.73</b>	746.87	1475.70	738.36	<b>1474.72</b>	737.86	9
6	<b>651.36</b>	326.19					G	<b>1361.69</b>	681.35	1344.66	672.84	<b>1343.68</b>	672.34	8
7	<b>764.45</b>	382.73					L	<b>1304.67</b>	652.84	1287.64	644.32	<b>1286.66</b>	643.83	7
8	<b>877.53</b>	439.27					L	<b>1191.59</b>	596.30	1174.56	587.78	<b>1173.57</b>	587.29	6
9	<b>1477.82</b>	739.41	1460.79	730.90			K	<b>1078.50</b>	539.75	1061.47	531.24	<b>1060.49</b>	530.75	5
10	<b>1564.85</b>	782.93	1547.83	774.42	1546.84	773.92	S	<b>478.21</b>	239.61	461.19	231.10	460.20	230.61	4
11	<b>1693.89</b>	847.45	1676.87	838.94	<b>1675.88</b>	838.45	E	391.18	196.09	374.16	187.58	<b>373.17</b>	187.09	3
12	<b>1808.92</b>	904.96	1791.89	896.45	<b>1790.91</b>	895.96	D	262.14	131.57	245.11	123.06	244.13	122.57	2
13							K	147.11	74.06	130.09	65.55			1

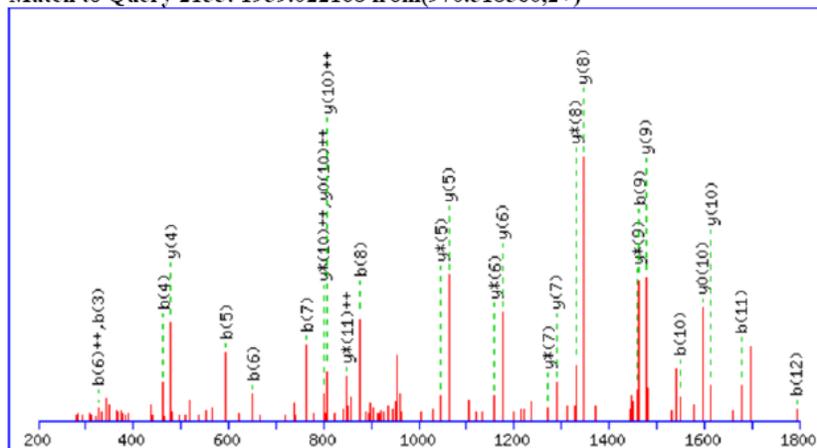
b)

MS/MS Fragmentation of **LLVHMGLLKSEDK**

Found in In\_vitro\_RanGap\_Sumo3\_withATP\_12March\_2010.raw - IPI:IPI00294879- Tax\_Id=9606

Gene\_Symbol=RANGAP1 Ran GTPase-activating protein 1

Match to Query 2155: 1939.022168 from(970.518360,2+)



Monoisotopic mass of neutral peptide Mr(calc): 1939.02 Variable modifications: K9 : Sumo 3 (K) Ions  
 Score: 63 Expect: 8.8e-05 Matches (**Bold Red**): 28/108 fragment ions using 46 most intense peaks

#	b	b <sup>++</sup>	b <sup>*</sup>	b <sup>*++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>*</sup>	y <sup>*++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	114.09	57.55					L							13
2	227.18	114.09					L	1826.94	913.98	1809.92	905.46	1808.93	904.97	12
3	<b>326.24</b>	163.63					V	1713.86	857.43	1696.83	<b>848.92</b>	1695.85	848.43	11
4	<b>463.30</b>	232.16					H	<b>1614.79</b>	<b>807.90</b>	1597.76	<b>799.39</b>	<b>1596.78</b>	<b>798.89</b>	10
5	<b>594.34</b>	297.68					M	<b>1477.73</b>	739.37	<b>1460.70</b>	730.86	1459.72	730.36	9
6	<b>651.36</b>	<b>326.19</b>					G	<b>1346.69</b>	673.85	<b>1329.66</b>	665.34	1328.68	664.84	8
7	<b>764.45</b>	382.73					L	<b>1289.67</b>	645.34	<b>1272.64</b>	636.83	1271.66	636.33	7
8	<b>877.53</b>	439.27					L	<b>1176.59</b>	588.80	<b>1159.56</b>	580.28	1158.57	579.79	6
9	<b>1462.82</b>	731.91	1445.79	723.40			K	<b>1063.50</b>	532.25	<b>1046.47</b>	523.74	1045.49	523.25	5
10	<b>1549.85</b>	775.43	1532.83	766.92	1531.84	766.42	S	<b>478.21</b>	239.61	461.19	231.10	460.20	230.61	4
11	<b>1678.89</b>	839.95	1661.87	831.44	1660.88	830.95	E	391.18	196.09	374.16	187.58	373.17	187.09	3
12	<b>1793.92</b>	897.46	1776.89	888.95	1775.91	888.46	D	262.14	131.57	245.11	123.06	244.13	122.57	2

Supplementary Figure 2.1: MS analyses of *in vitro* digestion products of GST-RanGAP1 (418-587) confirmed the SUMOylation site K524 for all SUMO paralogs. Tandem mass spectra of tryptic peptide LLVHMGLLKSEDK modified with (a) SUMO1, (b) SUMO3 mutants using CID activation

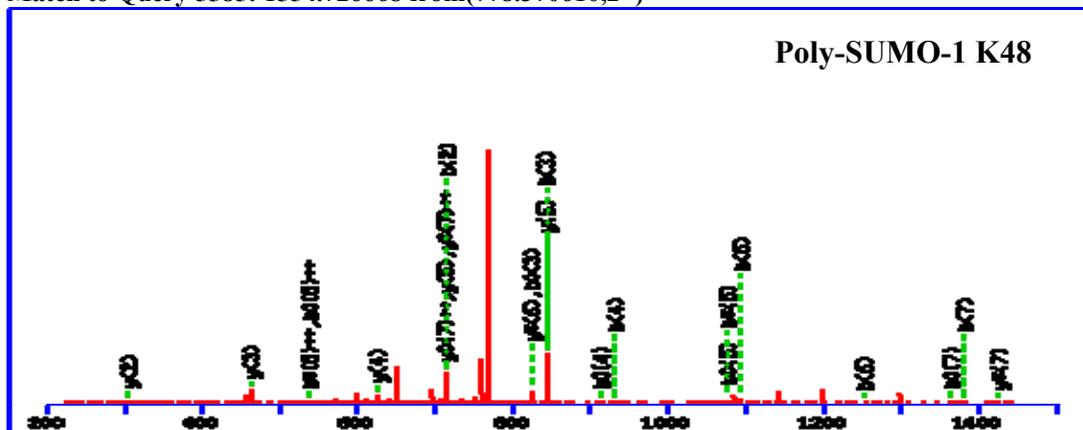
a)

MS/MS Fragmentation of **LKESYCQR**

Found in In\_vitro\_E25K\_Sumo1\_06April2010\_04.raw - IPI:IP100916532- Tax\_Id=9606

Gene\_Symbol=SUMO1 7 kDa protein

Match to Query 5585: 1554.726668 from(778.370610,2+)



Monoisotopic mass of neutral peptide Mr(calc): 1554.71 Variable modifications: K2 : SUMO EQTGG (K) (K) C6 : Carbamidomethyl (C) Ions Score: 21 Expect: 2.3 Matches (**Bold Red**): 22/70 fragment ions using 48 most intense peaks

#	b	b <sup>++</sup>	b*	b <sup>*++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>*++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	114.09	57.55					L							8
2	<b>714.38</b>	357.69	697.35	349.18			K	1442.63	721.82	<b>1425.61</b>	<b>713.31</b>	1424.62	<b>712.81</b>	7
3	<b>843.42</b>	422.21	826.39	413.70	<b>825.41</b>	413.21	E	<b>842.35</b>	421.68	<b>825.32</b>	413.16	824.34	412.67	6
4	<b>930.45</b>	465.73	913.43	457.22	<b>912.44</b>	456.72	S	<b>713.30</b>	357.16	696.28	348.64	695.29	348.15	5
5	<b>1093.52</b>	547.26	<b>1076.49</b>	<b>538.75</b>	<b>1075.51</b>	<b>538.26</b>	Y	<b>626.27</b>	313.64	609.24	305.13			4
6	<b>1253.55</b>	627.28	1236.52	618.76	1235.54	618.27	C	<b>463.21</b>	232.11	446.18	223.59			3
7	<b>1381.61</b>	691.31	1364.58	682.79	<b>1363.59</b>	682.30	Q	<b>303.18</b>	152.09	286.15	143.58			2
8							R	175.12	88.06	158.09	79.55			1

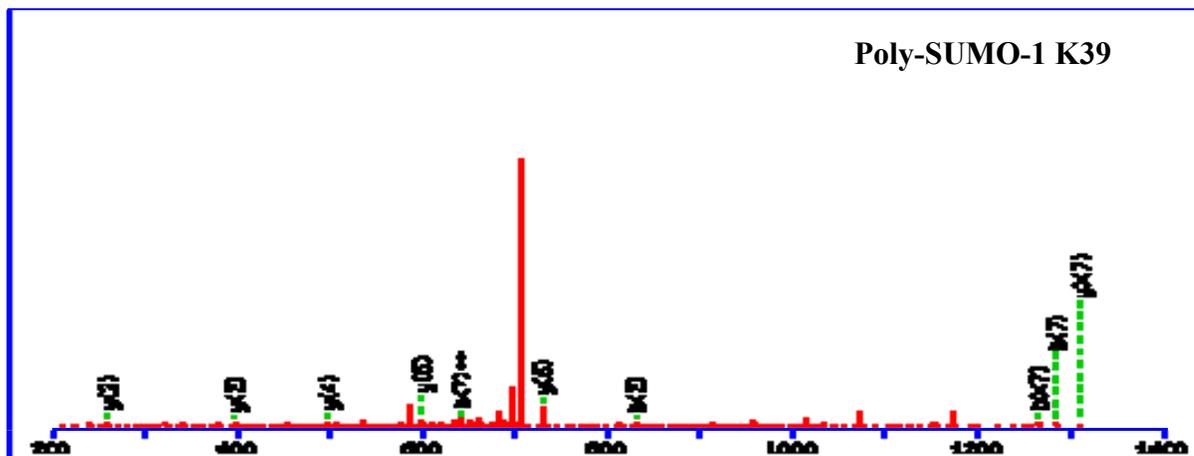


c)

MS/MS Fragmentation of **VKMTTHLK**

Found in In\_vitro\_E25K\_Sumo1\_06April2010\_06.raw -

Match to Query 5129: 1428.748068 from(715.381310,2+)



Monoisotopic mass of neutral peptide Mr(calc): 1428.74 Variable modifications: K2 : SUMO EQTGG

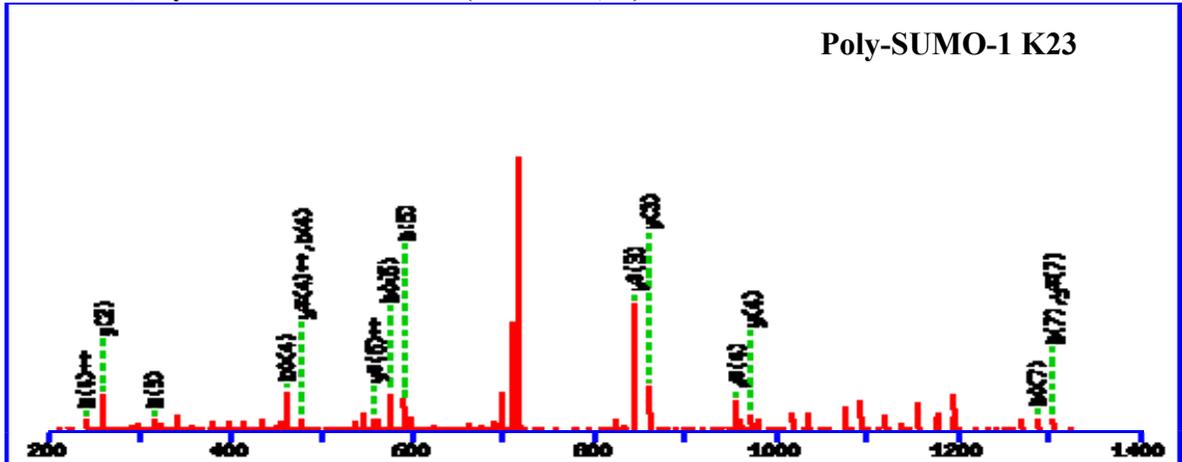
(K) (K) Ions Score: 23 Expect: 1.9 Matches (**Bold Red**): 10/70 fragment ions using 31 most intense peaks

#	<b>b</b>	<b>b<sup>++</sup></b>	<b>b<sup>*</sup></b>	<b>b<sup>*++</sup></b>	<b>b<sup>0</sup></b>	<b>b<sup>0++</sup></b>	Seq.	<b>y</b>	<b>y<sup>++</sup></b>	<b>y<sup>*</sup></b>	<b>y<sup>*++</sup></b>	<b>y<sup>0</sup></b>	<b>y<sup>0++</sup></b>	#
1	100.08	50.54					V							8
2	700.36	350.68	683.34	342.17			K	1330.68	665.84	1313.65	657.33	<b>1312.67</b>	656.84	7
3	<b>831.40</b>	416.21	814.38	407.69			M	<b>730.39</b>	365.70	713.37	357.19	712.38	356.69	6
4	932.45	466.73	915.42	458.22	914.44	457.72	T	<b>599.35</b>	300.18	582.32	291.67	581.34	291.17	5
5	1033.50	517.25	1016.47	508.74	1015.49	508.25	T	<b>498.30</b>	249.66	481.28	241.14	480.29	240.65	4
6	1170.56	585.78	1153.53	577.27	1152.55	576.78	H	<b>397.26</b>	199.13	380.23	190.62			3
7	<b>1283.64</b>	<b>642.32</b>	1266.61	633.81	<b>1265.63</b>	633.32	L	<b>260.20</b>	130.60	243.17	122.09			2
8							K	147.11	74.06	130.09	65.55			1

d)

MS/MS Fragmentation of **EGEYIKLK**  
 Found in In\_vitro\_E225K\_Sumo1\_06April2010\_04.raw -

Match to Query 5228: 1450.736088 from(726.375320,2+)

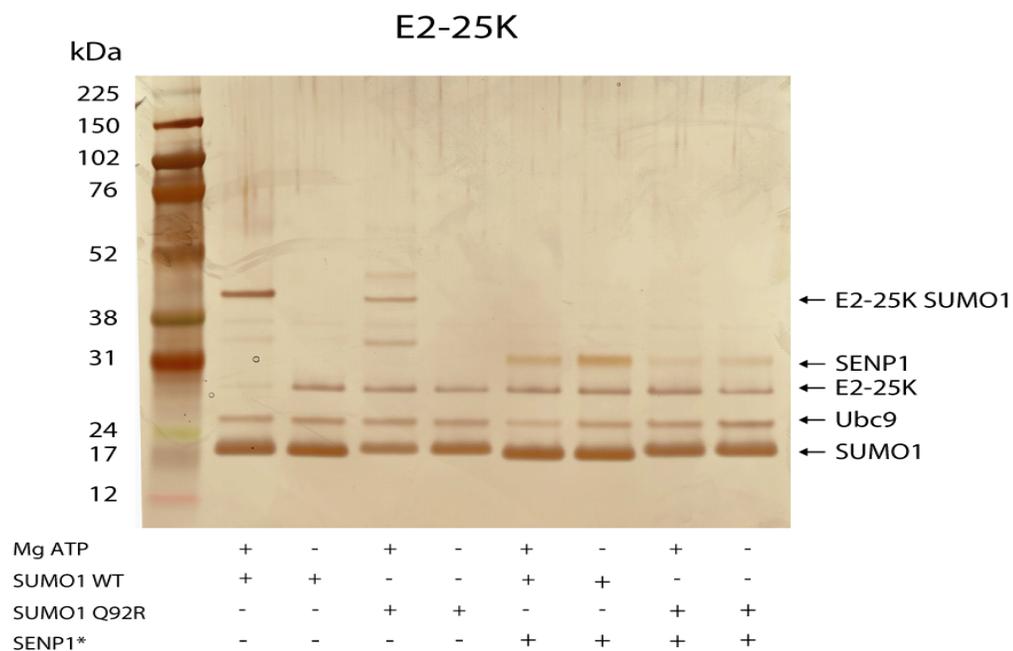


Monoisotopic mass of neutral peptide Mr(calc): 1450.73 Variable modifications: K6 : SUMO EQTGG  
 (K) (K) Ions Score: 11 Expect: 27 Matches (**Bold Red**): 16/64 fragment ions using 42 most intense peaks

#	b	b <sup>++</sup>	b*	b <sup>*++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>*++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	130.05	65.53			112.04	56.52	E							8
2	187.07	94.04			169.06	85.03	G	1322.70	661.85	<b>1305.67</b>	653.34	1304.68	652.85	7
3	<b>316.11</b>	158.56			298.10	149.56	E	1265.67	633.34	1248.65	624.83	1247.66	624.34	6
4	<b>479.18</b>	<b>240.09</b>			<b>461.17</b>	231.09	Y	1136.63	568.82	1119.60	<b>560.31</b>			5
5	<b>592.26</b>	296.63			<b>574.25</b>	287.63	I	<b>973.57</b>	487.29	<b>956.54</b>	<b>478.77</b>			4
6	1192.55	596.78	1175.52	588.26	1174.54	587.77	K	<b>860.48</b>	430.75	<b>843.46</b>	422.23			3
7	<b>1305.63</b>	653.32	1288.61	644.81	<b>1287.62</b>	644.31	L	<b>260.20</b>	130.60	243.17	122.09			2
8							K	147.11	74.06	130.09	65.55			1

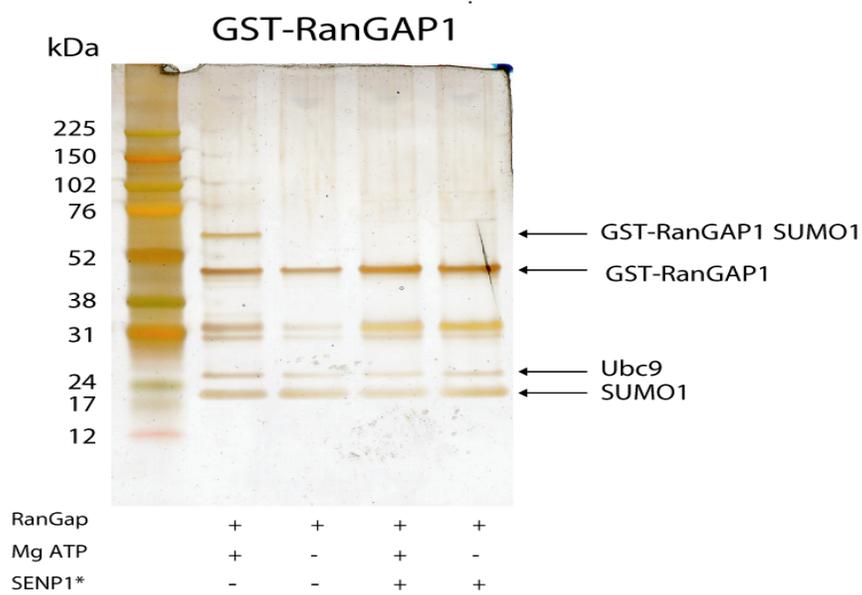
Supplementary Figure 2.2: MS analyses of *in vitro* SUMOylation enabled the identification of sites of polySUMOylation. Tandem mass spectra (CID) of tryptic peptides from SUMO1 mutant showing cross-linkage at (a) K48, (b) K37, (c) K39 and (d) K23 residues. PolySUMOylation was observed from *in vitro* SUMOylation of E2-ligase and RanGAP1 with His<sub>6</sub>-SUMO1 mutant.

a)



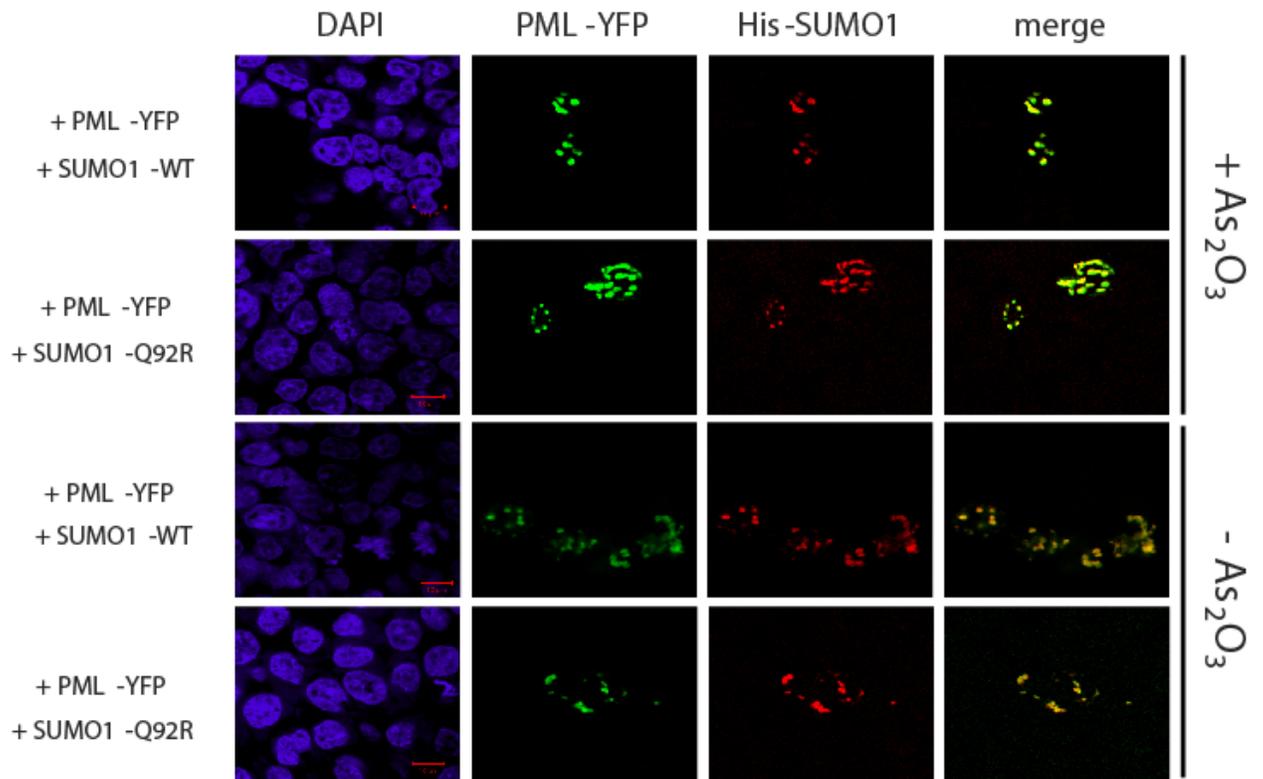
\*SUMOylation reactions were stopped with EDTA before adding SENP1

b)

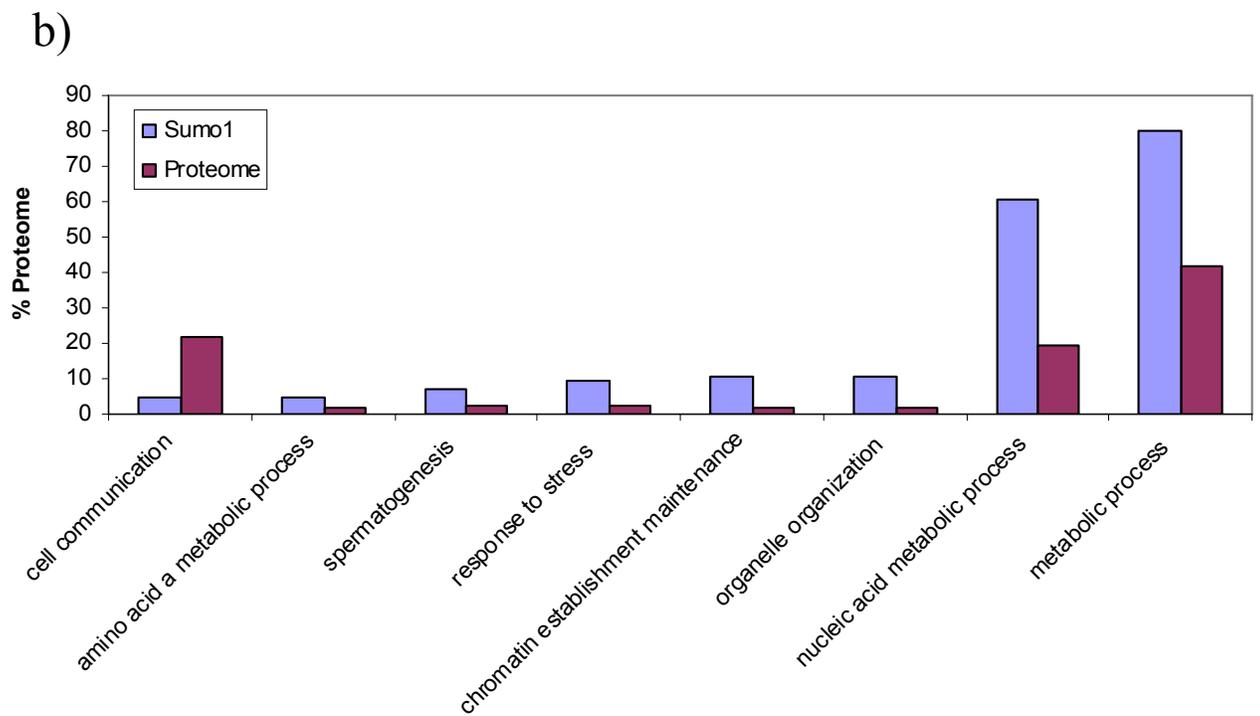
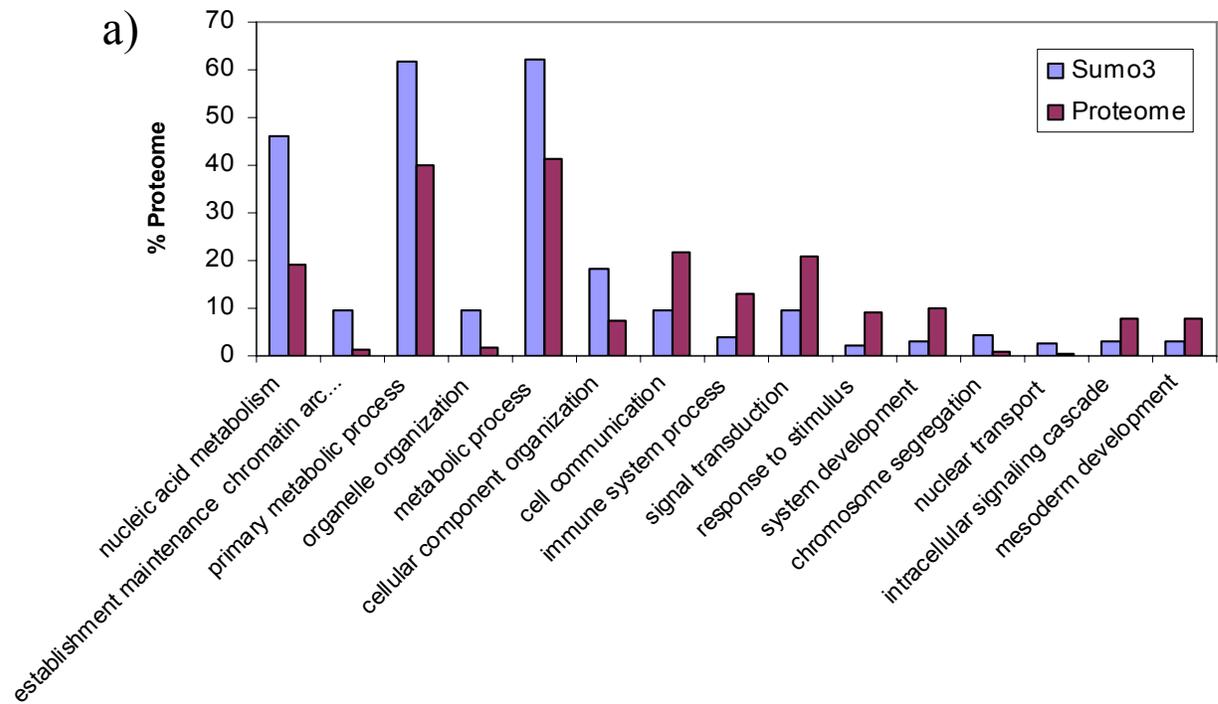


\*Sumoylation reaction were stopped with EDTA before adding SENP1

Supplementary Figure 2.3 : *In vitro* deSUMOylation assay of E2-ligase and RanGAP1 using SENP1. a) Silver stained gel of reaction products from E2-25 (deSUMOylation by SENP1 following *in vitro* SUMOylation by Ubc9. b) Silver stained gel of reaction products from GST-RanGAP1 (418-587) deSUMOylation by SENP1 following *in vitro* SUMOylation by Ubc9. c) Silver stained gel showing the comparison deSUMOylation of WT and His6-SUMO1 mutant by SENP1. d) Anti-His immunoblot of SUMOylated E2-25 K with and without SENP1. e) Anti GST immunoblot of SUMOylated GST-PML (485-495). Approximately 1-2 ug of protein substrate was used in each experiment. *In vitro* SUMOylation conditions as indicated in the experimental section. Following the *in vitro* SUMOylation, the SAE1/2 enzyme was inhibited with EDTA (20 mM) prior to incubation with 1  $\mu$ M SENP1 (Boston Biochem) at 37°C for 2h.

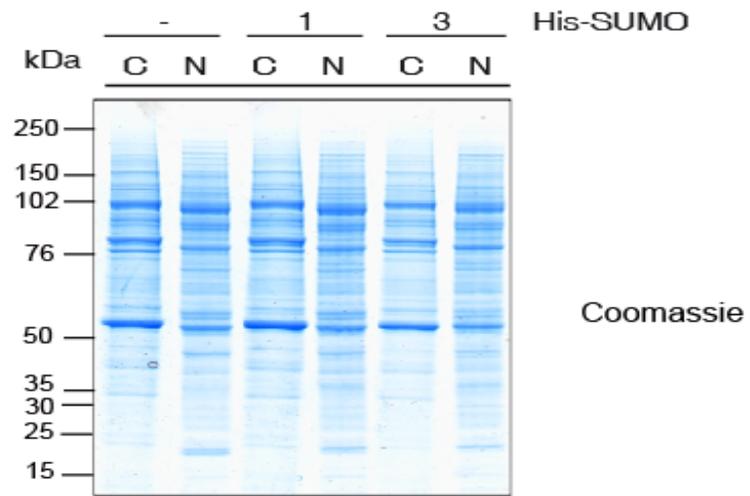


Supplementary Figure 2.4 : Comparison of His-SUMO1 WT and His-SUMO1 mutant to SUMOylate PML and to colocalize within nuclear bodies. Immunofluorescence of HEK293 cells co-transfected with YFP-PML and His-SUMO1 WT or mutant revealed using anti-His antibody. As<sub>2</sub>O<sub>3</sub> induced aggregation of SUMOylated PML in NBs for both His<sub>6</sub>-SUMO1 WT and mutant compared to untreated cells.

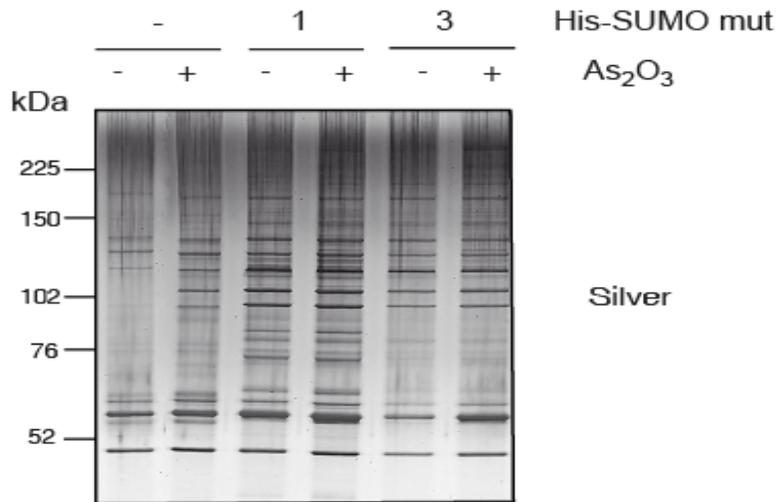


Supplementary Figure 2.5: Analysis of Gene Ontology terms associated with biological processes from a) SUMO-3 and b) SUMO-1 protein candidates using the software PANTHER (see Annex IV, Supplementary Tables 1 and 2 for protein list).

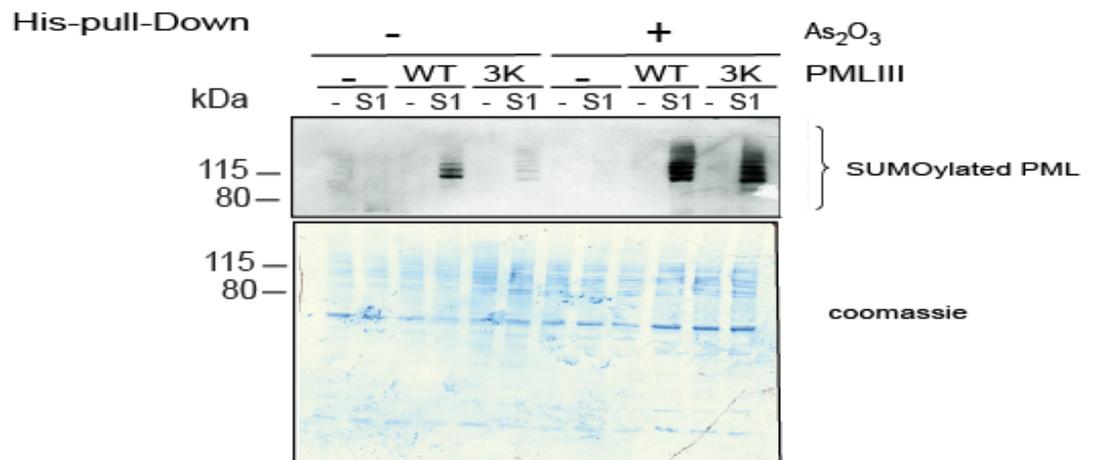
a)



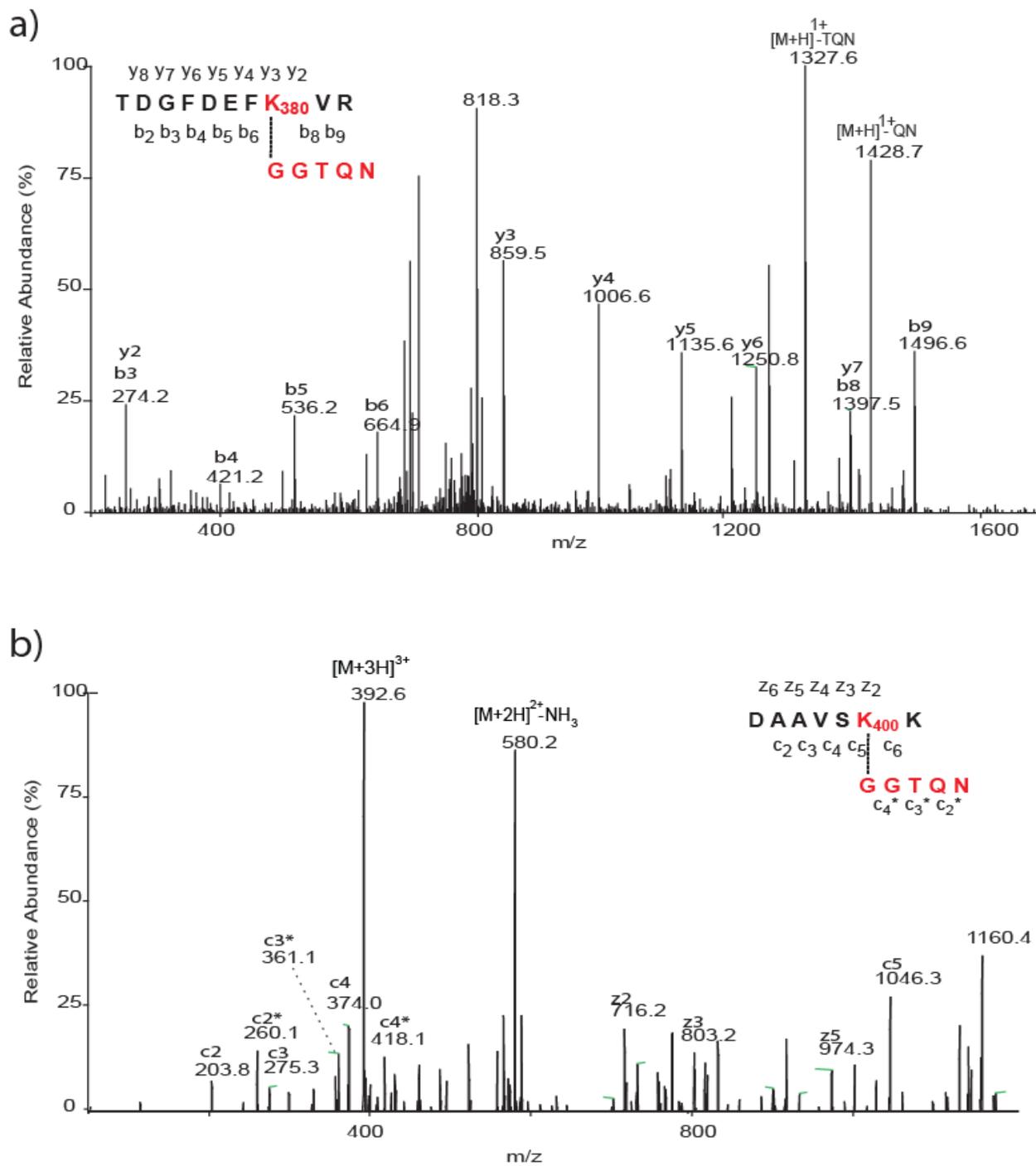
b)



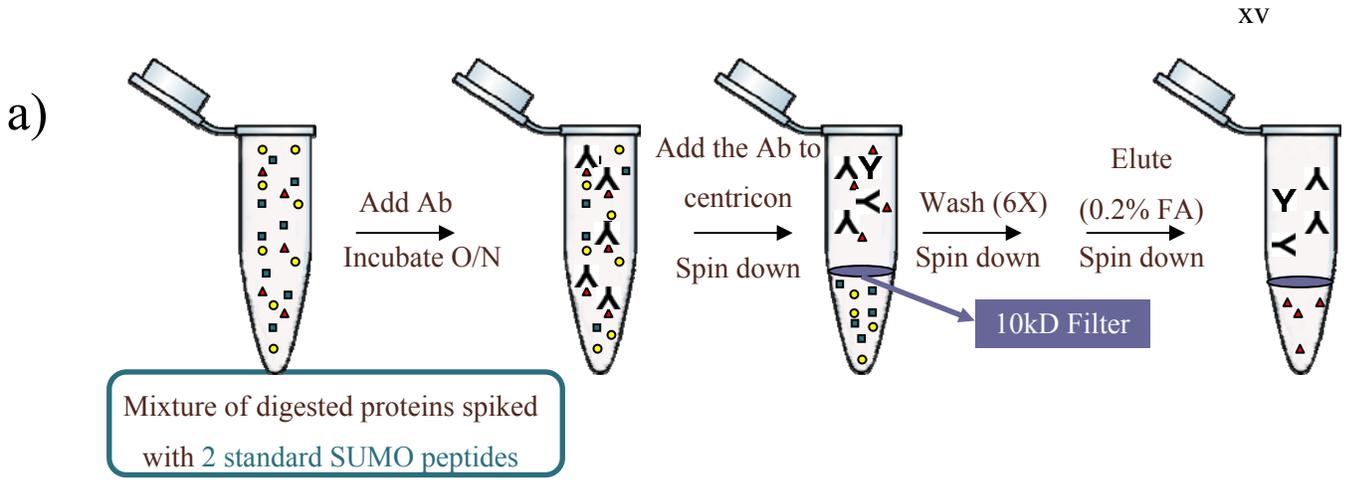
c)



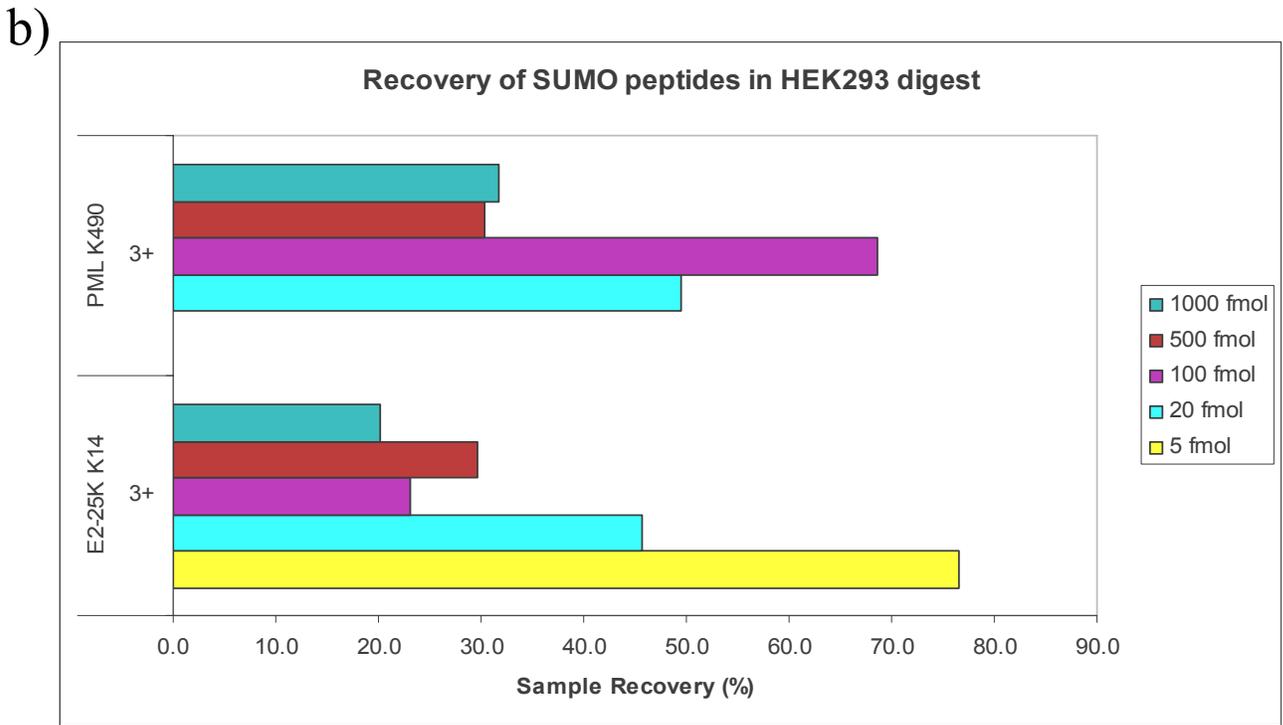
Supplementary Figure 2.6: NTA purified extracts from control HEK293 and HEK293-His-SUMO cells. a) Loading control (Fig. 4a) for the comparison of cytosol (C) and nuclear (N) extracts revealed using silver staining. b) Loading control for Fig, 4b using silver staining. c) Immunoblot (PML) and coomassie stained gel of NTA purified extracts from control HEK293 and HEK293-His-SUMO1 cells.



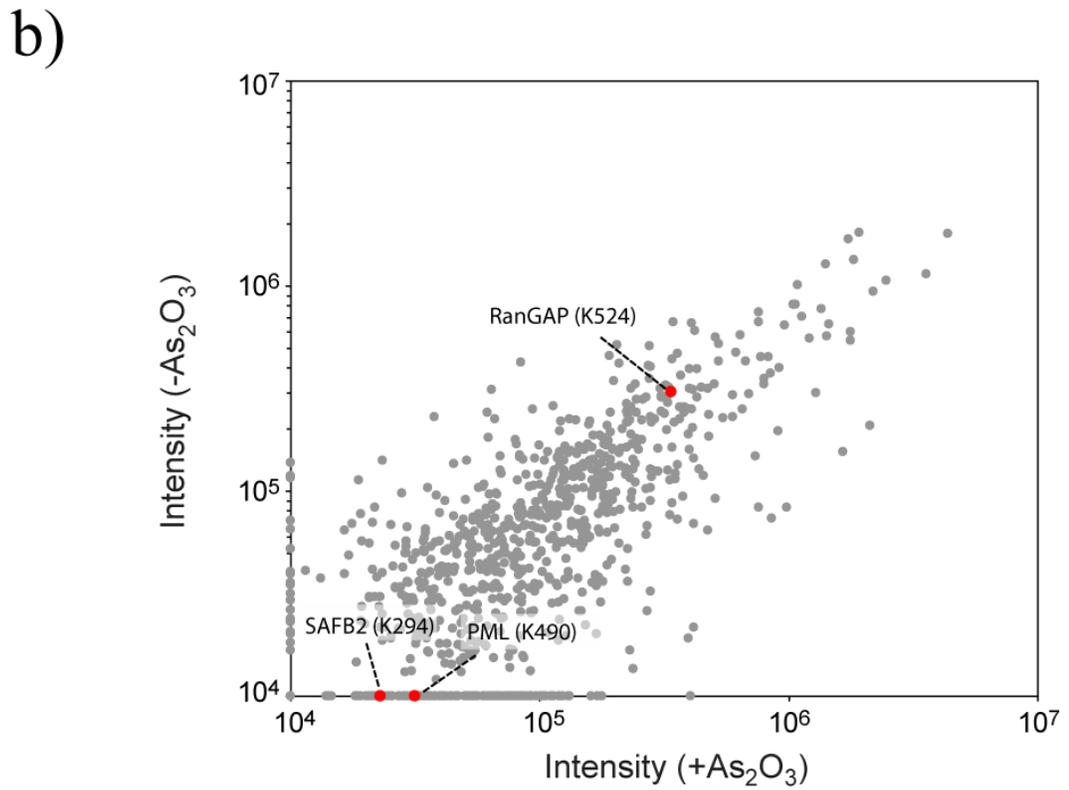
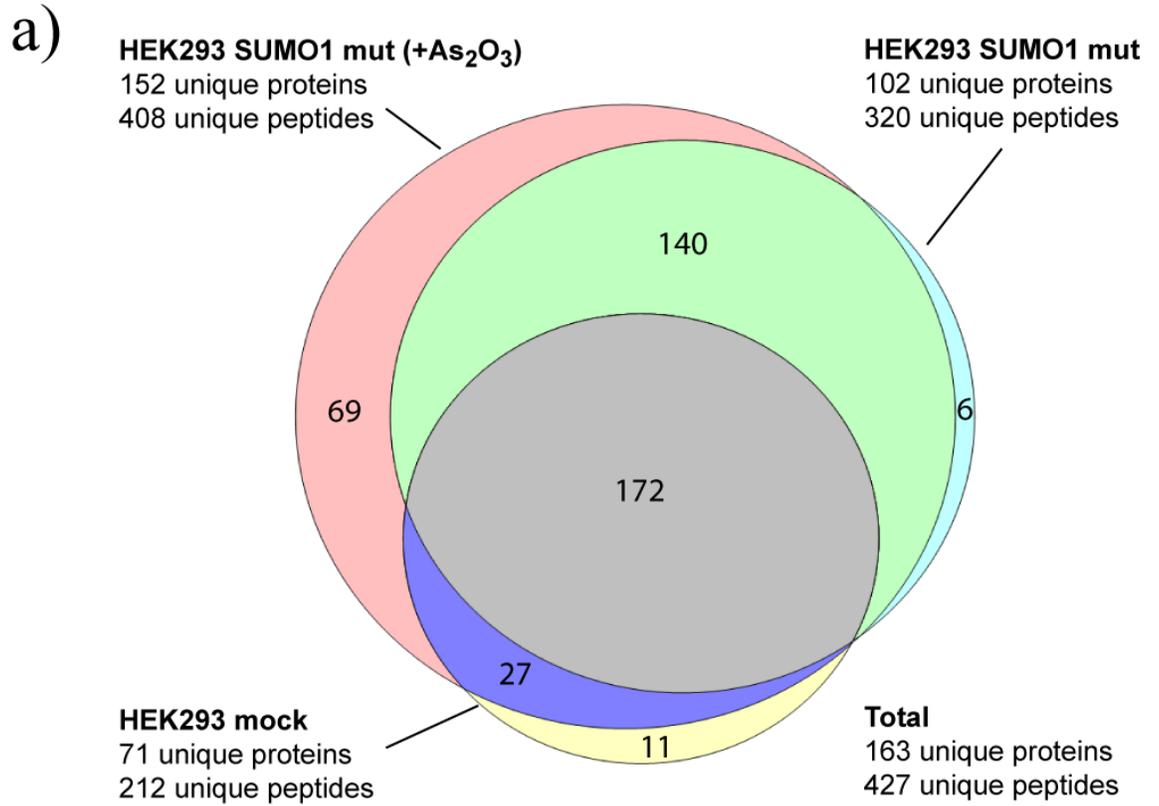
Supplementary Figure 2.7: Tandem mass spectra of synthetic peptides to confirm SUMOylation of sites K380 and K400. MS/MS spectra of precursor ions (a)  $m/z$  835.89<sup>2+</sup> and (b)  $m/z$  392.54<sup>3+</sup> obtained using CID and ETD activation, respectively.



-  Anti-GGTQE antibody
-  SUMOylated peptides
-  Other peptides



Supplementary Figure 2.8 : (a)Design of an immunoprecipitation experiment using a centrifugal filter device. Sumoylated peptides are specifically retained in the upper chamber. (b)Recovery yields of SUMOylated peptides EFK\*EVLK (SUMOylated K14 from E2-25K) and KVIK\*MESEEGKEAR (SUMOylated K490 from E2-25K) spiked at different levels in a HEK293 tryptic digest. \* indicates site of attachment of GGTQE on modified Lys residue.



Supplementary Figure 2.9: Mass spectrometry analyses of SUMOylated proteins from Immunoaffinity purified NTA extracts of mock HEK293 and HEK293 His<sub>6</sub>-SUMO1 in As<sub>2</sub>O<sub>3</sub>-stimulated and non-stimulated cells. a) Venn diagram showing the overlapping distribution of peptides in each cell extract. Proteins identified with at least 2 peptides were considered for the comparison. b) Intensity distribution of peptide ions identified in the HEK293 His<sub>6</sub>-SUMO3 in As<sub>2</sub>O<sub>3</sub>-stimulated and non-stimulated cells. The change in abundance of tryptic peptides from PML (K490), RanGAP1 (K524) and SAFB2 (K294) are indicated on the scatter plot.

**Annex II: Supplementary Figure 2.10 for Chapter 2**  
**(A Novel Proteomics Approach to Identify SUMOylated Proteins and**  
**their Modification Sites in Human Cells)**

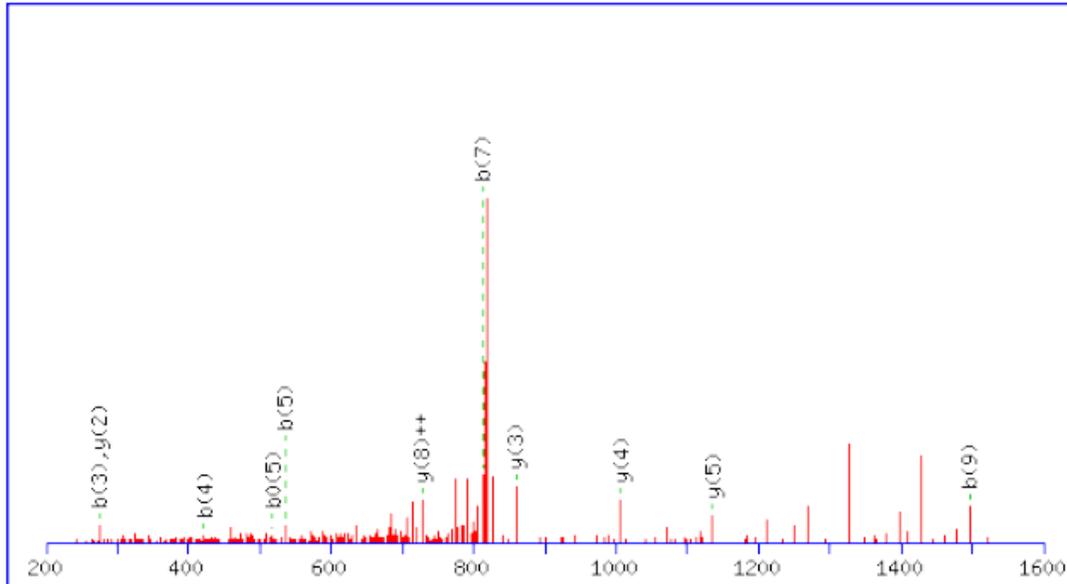
Supplementary Figure 2.10: Tandem Mass Spectra of All Identified SUMO-3 Peptides Following Database Search and Manual Validation.

## Peptide No.1

MS/MS Fragmentation of **TDGFDEFKVR**

Found in Sumo3\_with\_AS\_21Nov2010\_1.RAW - **IPI00876842**- Tax\_Id=9606 Gene\_Symbol=- PML-RAR protein

Match to Query 7269: 1669.786748 from(835.900650,2+)



Monoisotopic mass of neutral peptide Mr(calc): 1669.77 Variable modifications: K8 : Sumo 3 (K) Ions Score: 26

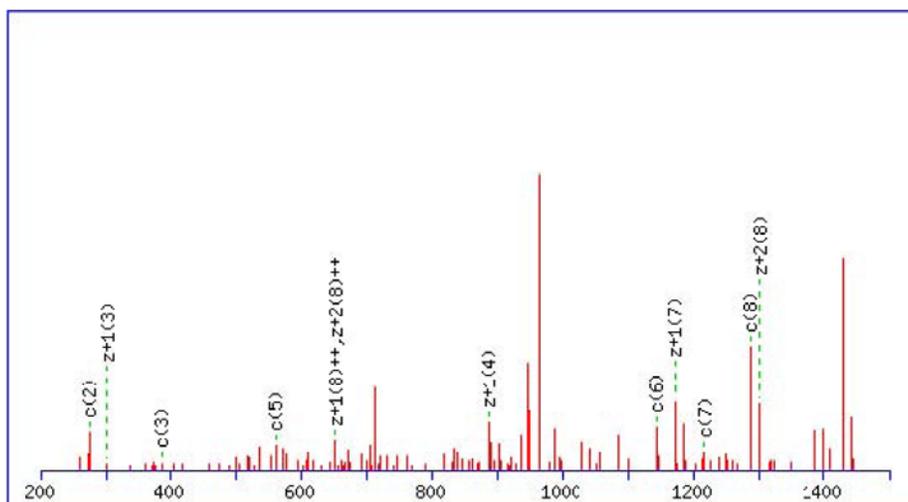
Expect: 1.1 Matches (**Bold Red**): 11/86 fragment ions using 31 most intense peaks

#	b	b <sup>++</sup>	b <sup>*</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>*</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	102.05	51.53			84.04	42.53	T							10
2	217.08	109.04			199.07	100.04	D	1569.73	785.37	1552.70	776.85	1551.72	776.36	9
3	<b>274.10</b>	137.56			256.09	128.55	G	1454.70	<b>727.85</b>	1437.68	719.34	1436.69	718.85	8
4	<b>421.17</b>	211.09			403.16	202.08	F	1397.68	699.34	1380.65	690.83	1379.67	690.34	7
5	<b>536.20</b>	268.60			<b>518.19</b>	259.60	D	1250.61	625.81	1233.59	617.30	1232.60	616.80	6
6	665.24	333.12			647.23	324.12	E	<b>1135.59</b>	568.30	1118.56	559.78	1117.57	559.29	5
7	<b>812.31</b>	406.66			794.30	397.65	F	<b>1006.54</b>	503.78	989.52	495.26			4
8	1397.60	699.30	1380.57	690.79	1379.59	690.30	K	<b>859.47</b>	430.24	842.45	421.73			3
9	<b>1496.67</b>	748.84	1479.64	740.32	1478.65	739.83	V	<b>274.19</b>	137.60	257.16	129.08			2
10							R	175.12	88.06	158.09	79.55			1

## Peptide No.2

MS/MS Fragmentation of **KQLATKAAR** Found in  
Sumo3\_wo\_AS\_21Nov2010\_4.RAW -**IPI00171611**- Tax\_Id=9606  
Gene\_Symbol=HIST2H3A;HIST2H3D; HIST2H3C Histone H3.2

Match to Query 2256: 1442.797152 from(481.939660,3+)



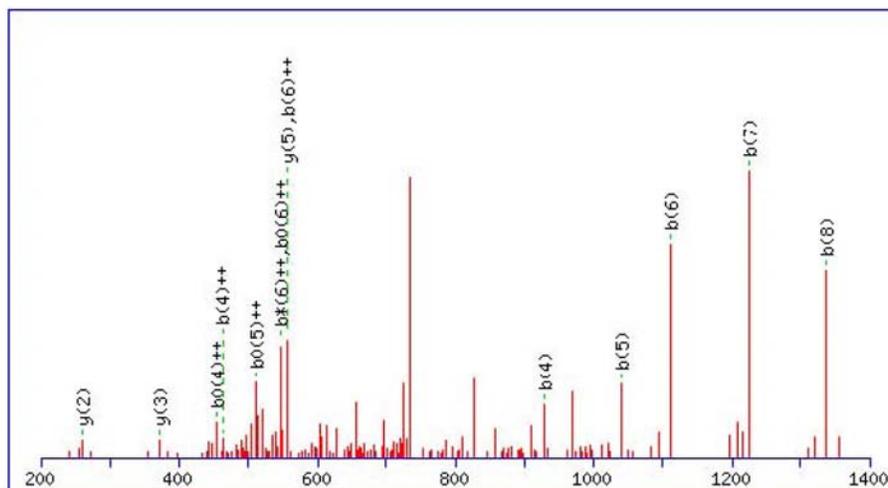
Monoisotopic mass of neutral peptide Mr(calc): 1442.80 Variable modifications: K6 :  
Sumo 3 (K) Ions Score: 13 Expect: 19 Matches (**Bold Red**): 12/48 fragment ions using 38  
most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	146.13	73.57	K					9
2	<b>274.19</b>	137.60	Q	1299.69	<b>650.35</b>	<b>1300.70</b>	<b>650.85</b>	8
3	<b>387.27</b>	194.14	L	<b>1171.63</b>	586.32	1172.64	586.82	7
4	458.31	229.66	A	1058.55	529.78	1059.55	530.28	6
5	<b>559.36</b>	280.18	T	987.51	494.26	988.52	494.76	5
6	<b>1144.64</b>	572.83	K	<b>886.46</b>	443.73	887.47	444.24	4
7	<b>1215.68</b>	608.34	A	<b>301.17</b>	151.09	302.18	151.59	3
8	<b>1286.72</b>	643.86	A	230.14	115.57	231.15	116.08	2
9			R	159.10	80.05	160.11	80.56	1

### Peptide No.3

MS/MS Fragmentation of **REKGLALLK** Found in  
Sumo3\_with\_AS\_21Nov2010\_5.RAW -IPI00008456- Tax\_Id=9606 Gene\_Symbol=HSF4  
Isoform HSF4B of Heat shock factor protein 4

Match to Query 6025: 1483.839728 from(742.927140,2+)



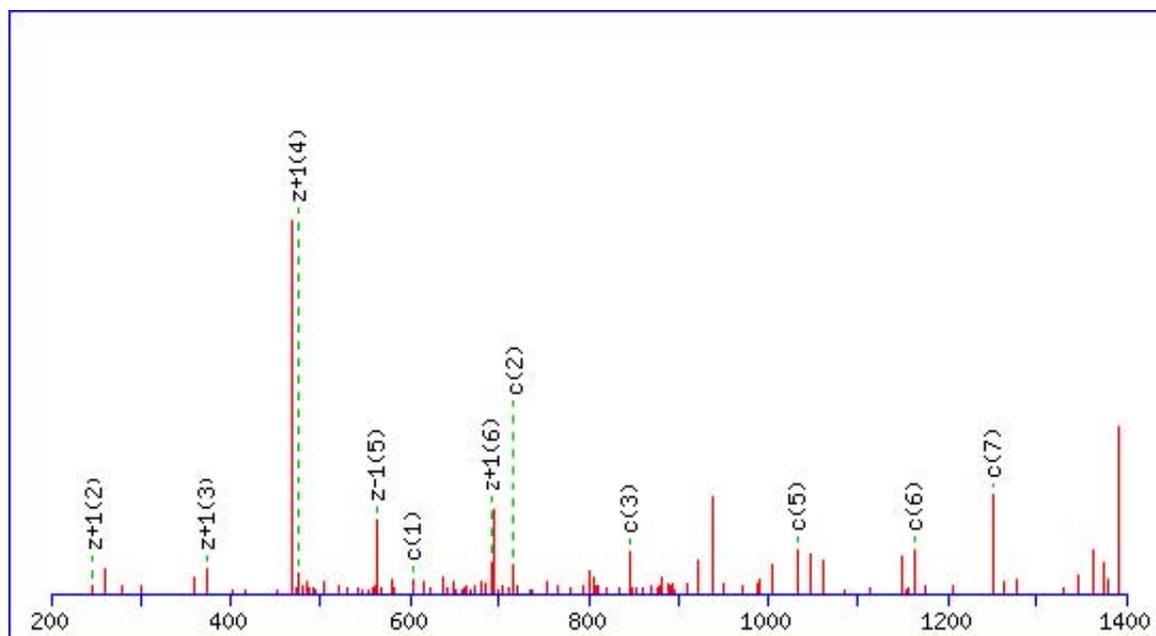
Monoisotopic mass of neutral peptide Mr(calc): 1483.85 Variable modifications: K3 :  
Sumo 3 (K) Ions Score: 38 Expect: 0.037 Matches (**Bold Red**): 14/80 fragment ions using  
19 most intense peaks

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	157.11	79.06	140.08	70.54			R							9
2	286.15	143.58	269.12	135.07	268.14	134.57	E	1328.75	664.88	1311.73	656.37	1310.74	655.87	8
3	871.44	436.22	854.41	427.71	853.43	427.22	K	1199.71	600.36	1182.68	591.85			7
4	<b>928.46</b>	<b>464.73</b>	911.43	456.22	910.45	<b>455.73</b>	G	614.42	307.72	597.40	299.20			6
5	<b>1041.54</b>	521.28	1024.52	512.76	1023.53	<b>512.27</b>	L	<b>557.40</b>	279.20	540.38	270.69			5
6	<b>1112.58</b>	<b>556.79</b>	1095.55	<b>548.28</b>	1094.57	<b>547.79</b>	A	444.32	222.66	427.29	214.15			4
7	<b>1225.66</b>	613.34	1208.64	604.82	1207.65	604.33	L	<b>373.28</b>	187.14	356.25	178.63			3
8	<b>1338.75</b>	669.88	1321.72	661.36	1320.74	660.87	L	<b>260.20</b>	130.60	243.17	122.09			2
9							K	147.11	74.06	130.09	65.55			1

### Peptide No.4

MS/MS Fragmentation of **KLESTESR** Found in  
Sumo3\_with\_AS\_21Nov2010\_2.RAW -IPI00514817- Tax\_Id=9606  
Gene\_Symbol=LMNA Lamin A/C

Match to Query 2258: 1405.678032 from(469.566620,3+)



Monoisotopic mass of neutral peptide Mr(calc): 1405.68 Variable modifications: K1 :  
Sumo 3 (K) Ions Score: 19 Expect: 8 Matches (**Bold Red**): 11/42 fragment ions using 37  
most intense peaks

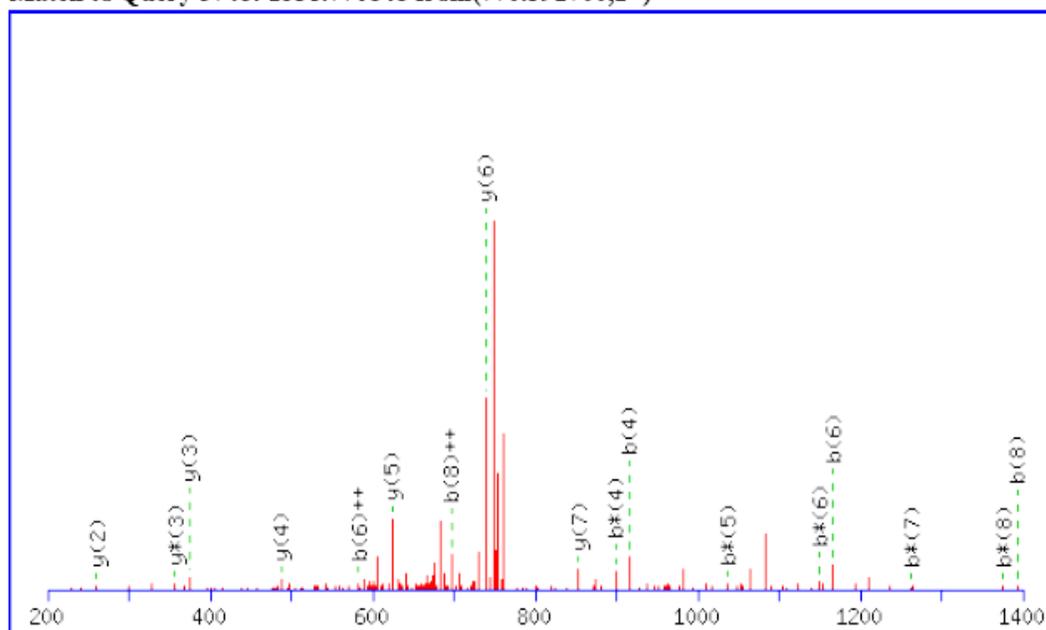
#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	<b>603.32</b>	302.16	<b>K</b>					8
2	<b>716.40</b>	358.71	<b>L</b>	805.38	403.19	806.39	403.70	7
3	<b>845.45</b>	423.23	<b>E</b>	<b>692.30</b>	346.65	693.30	347.16	6
4	932.48	466.74	<b>S</b>	<b>563.25</b>	282.13	564.26	282.63	5
5	<b>1033.53</b>	517.27	<b>T</b>	<b>476.22</b>	238.61	477.23	239.12	4
6	<b>1162.57</b>	581.79	<b>E</b>	<b>375.17</b>	188.09	376.18	188.59	3
7	<b>1249.60</b>	625.30	<b>S</b>	<b>246.13</b>	123.57	247.14	124.07	2
8			<b>R</b>	159.10	80.05	160.11	80.56	1

## Peptide No.5

MS/MS Fragmentation of **TKNNHINLK**

Found in Sumo3\_wo\_AS\_21Nov2010\_2.RAW - **IPI00888867**- Tax\_Id=9606 Gene\_Symbol=LOC100130966 similar to SMT3 suppressor of mif two 3 homolog 2

Match to Query 5748: 1538.770848 from(770.392700,2+)



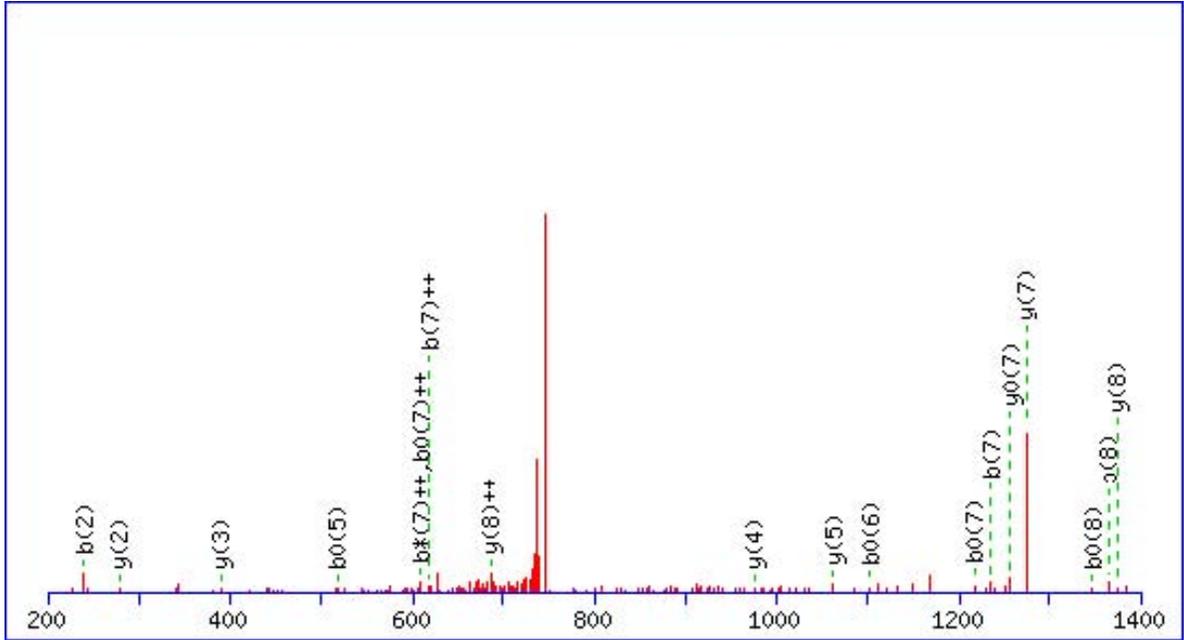
Monoisotopic mass of neutral peptide Mr(calc): 1538.78 Variable modifications: K2 : Sumo 3 (K) N4 : Deamidated (NQ) Ions Score: 23 Expect: 2.4 Matches (**Bold Red**): 17/78 fragment ions using 38 most intense peaks

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	#
1	102.05	51.53			84.04	42.53	T					9
2	687.34	344.17	670.32	335.66	669.33	335.17	K	1438.74	719.87	1421.71	711.36	8
3	801.38	401.20	784.36	392.68	783.37	392.19	N	<b>853.45</b>	427.23	836.43	418.72	7
4	<b>916.41</b>	458.71	<b>899.39</b>	450.20	898.40	449.70	N	<b>739.41</b>	370.21	722.38	361.70	6
5	1053.47	527.24	<b>1036.44</b>	518.73	1035.46	518.23	H	<b>624.38</b>	312.70	607.36	304.18	5
6	<b>1166.55</b>	<b>583.78</b>	<b>1149.53</b>	575.27	1148.54	574.78	I	<b>487.32</b>	244.17	470.30	235.65	4
7	1280.60	640.80	<b>1263.57</b>	632.29	1262.59	631.80	N	<b>374.24</b>	187.62	<b>357.21</b>	179.11	3
8	<b>1393.68</b>	<b>697.34</b>	<b>1376.66</b>	688.83	1375.67	688.34	L	<b>260.20</b>	130.60	243.17	122.09	2
9							K	147.11	74.06	130.09	65.55	1

### Peptide No.6

MS/MS Fragmentation of **HTPLSKLMK** Found in  
Sumo3\_with\_AS\_21Nov2010\_5.RAW -IPI00455745- Tax\_Id=9606  
Gene\_Symbol=LOC100133788 similar to SMT3B protein

Match to Query 6221: 1510.788848 from(756.401700,2+)



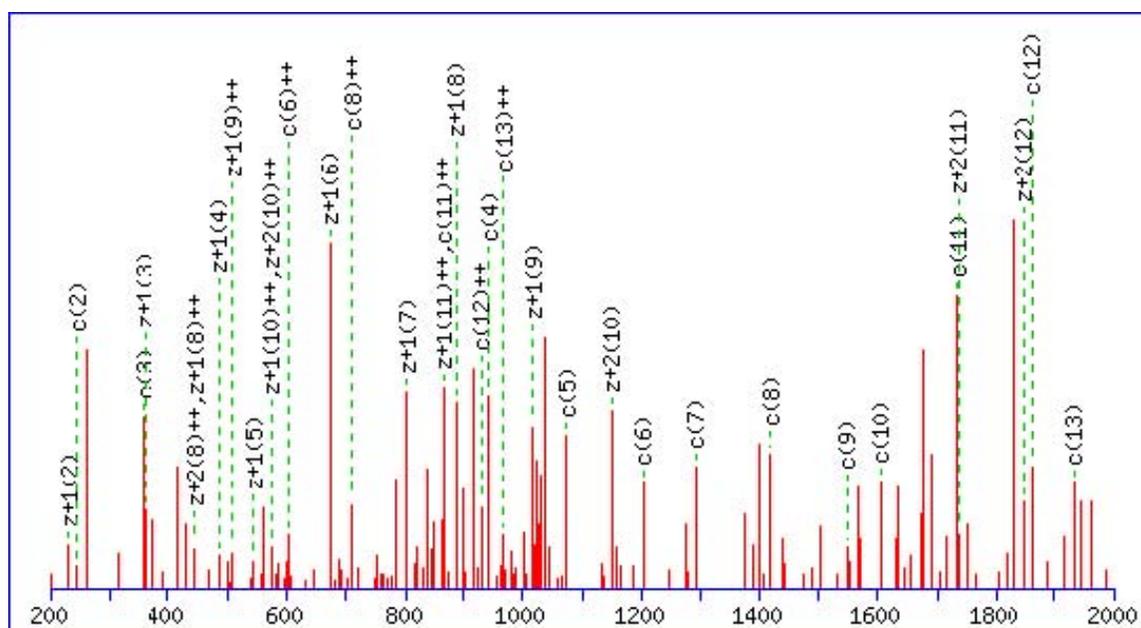
Monoisotopic mass of neutral peptide Mr(calc): 1510.79 Variable modifications: K6 :  
Sumo 3 (K) Ions Score: 20 Expect: 2.7 Matches (**Bold Red**): 18/76 fragment ions using 50  
most intense peaks

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	128.07	69.54					H							9
2	<b>239.11</b>	120.06			221.10	111.06	T	<b>1374.74</b>	<b>687.87</b>	1357.71	679.36	1356.73	678.87	8
3	336.17	168.59			318.16	159.58	P	<b>1273.69</b>	637.35	1256.67	628.84	<b>1255.68</b>	628.34	7
4	449.25	225.13			431.24	216.12	L	1176.64	588.82	1159.61	580.31	1158.63	579.82	6
5	526.28	268.64			<b>518.27</b>	259.64	S	<b>1063.56</b>	532.28	1046.52	523.77	1045.55	523.28	5
6	1121.57	561.29	1104.54	552.78	<b>1103.56</b>	552.28	K	<b>976.52</b>	483.77	959.50	480.25			4
7	<b>1234.65</b>	<b>617.83</b>	1217.63	<b>609.32</b>	<b>1216.64</b>	<b>608.83</b>	L	<b>391.24</b>	195.12	374.21	187.61			3
8	<b>1365.69</b>	683.35	1348.67	674.84	<b>1347.68</b>	674.35	M	<b>278.15</b>	139.58	261.12	131.07			2
9							K	147.11	74.06	130.05	65.55			1

## Peptide No.7

MS/MS Fragmentation of **KVIKMESEEGKEAR** Found in  
Sumo3\_wo\_AS\_21Nov2010\_4.RAW -**IPI00022348**- Tax\_Id=9606 Gene\_Symbol=PML  
Isoform PML-1 of Probable transcription factor PML

Match to Query 4747: 2090.034976 from(523.516020,4+)



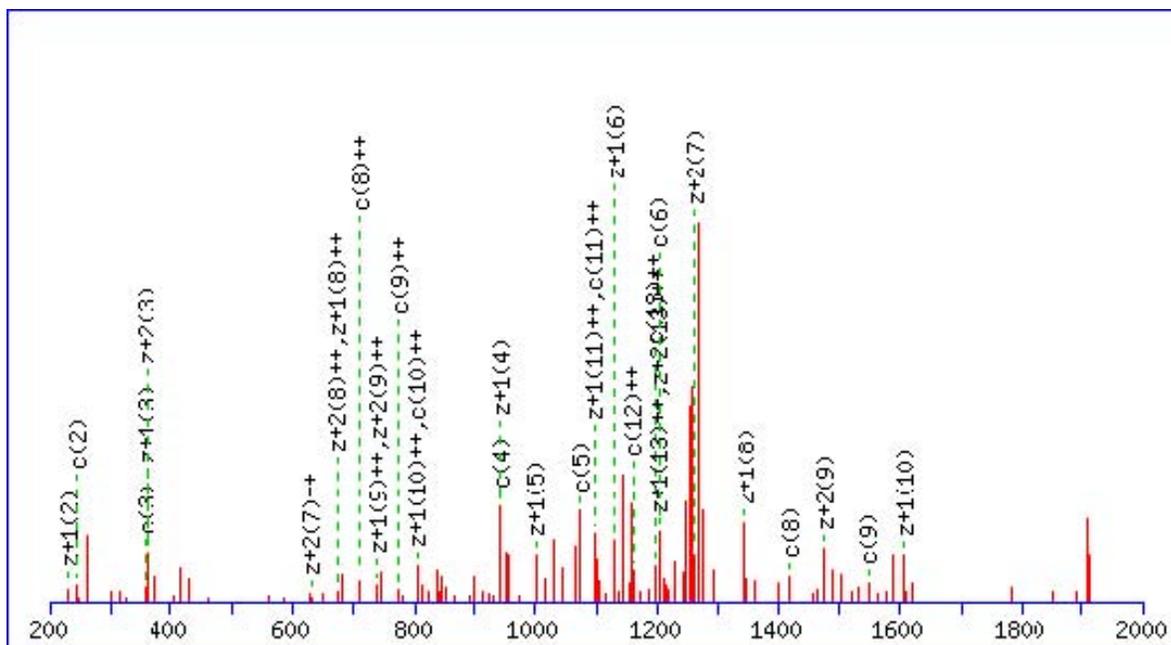
Monoisotopic mass of neutral peptide Mr(calc): 2090.04 Variable modifications: K4 :  
Sumo 3 (K) Ions Score: 60 Expect: 0.001 Matches (**Bold Red**): 34/78 fragment ions using  
71 most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	146.13	73.57	K					14
2	<b>245.20</b>	123.10	V	1946.94	973.97	1947.94	974.48	13
3	<b>358.28</b>	179.64	I	1847.87	924.44	<b>1848.88</b>	924.94	12
4	<b>943.57</b>	472.29	K	1734.78	<b>867.90</b>	<b>1735.79</b>	868.40	11
5	<b>1074.61</b>	537.81	M	1149.50	<b>575.25</b>	<b>1150.50</b>	<b>575.76</b>	10
6	<b>1203.65</b>	<b>602.33</b>	E	<b>1018.46</b>	<b>509.73</b>	1019.46	510.24	9
7	<b>1290.68</b>	645.85	S	<b>889.41</b>	<b>445.21</b>	890.42	<b>445.71</b>	8
8	<b>1419.73</b>	<b>710.37</b>	E	<b>802.38</b>	401.69	803.39	402.20	7
9	<b>1548.77</b>	774.89	E	<b>673.34</b>	337.17	674.35	337.68	6
10	<b>1605.79</b>	803.40	G	<b>544.30</b>	272.65	545.30	273.16	5
11	<b>1733.89</b>	<b>867.45</b>	K	<b>487.27</b>	244.14	488.28	244.64	4
12	<b>1862.93</b>	<b>931.97</b>	E	<b>359.18</b>	180.09	360.19	180.60	3
13	<b>1933.96</b>	<b>967.49</b>	A	<b>230.14</b>	115.57	231.15	116.08	2
14			R	159.10	80.05	160.11	80.56	1

## Peptide No.8

MS/MS Fragmentation of **KVIKMESEEGKEAR** Found in  
Sumo3\_with\_AS\_21Nov2010\_4.RAW -IPI00303999- Tax\_Id=9606 Gene\_Symbol=PML  
Isoform PML-2 of Probable transcription factor PML

Match to Query 5504: 2547.235856 from(637.816240,4+)



Monoisotopic mass of neutral peptide Mr(calc): 2547.23 Variable modifications: K4 :  
Sumo 3 (K) K11 : Sumo 3

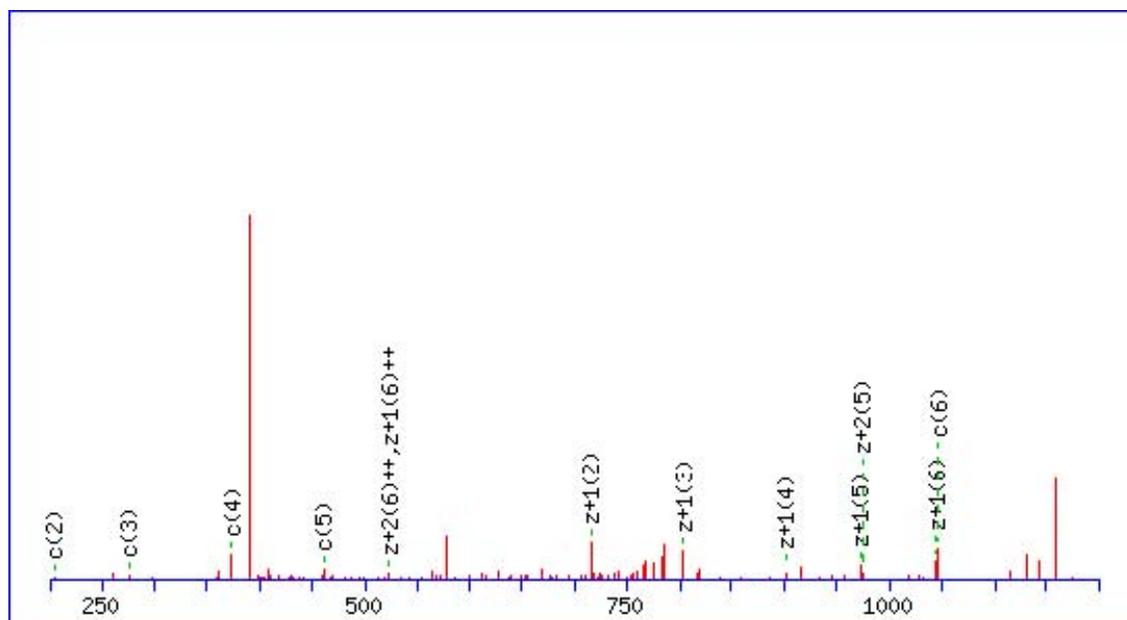
(K) Ions Score: 16 Expect: 34 Matches (**Bold Red**): 32/78 fragment ions using 92 most  
intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	146.13	73.57	K					14
2	245.20	123.10	V	2404.13	1202.57	2405.14	1203.07	13
3	358.28	179.64	I	2305.06	1153.03	2306.07	1153.54	12
4	943.57	472.29	K	2191.98	1096.49	2192.98	1097.00	11
5	1074.61	537.81	M	1606.69	803.85	1607.70	804.35	10
6	1203.65	602.33	E	1475.65	738.33	1476.66	738.83	9
7	1290.68	645.85	S	1346.61	673.81	1347.61	674.31	8
8	1419.73	710.37	E	1259.57	630.29	1260.58	630.79	7
9	1548.77	774.89	E	1130.53	565.77	1131.54	566.27	6
10	1605.79	803.40	G	1001.49	501.25	1002.50	501.75	5
11	2191.08	1096.04	K	944.47	472.74	945.47	473.24	4
12	2320.12	1160.56	E	359.18	180.09	360.19	180.60	3
13	2391.16	1196.08	A	230.14	115.57	231.15	116.08	2
14			R	159.10	80.05	160.11	80.56	1

## Peptide No.9

MS/MS Fragmentation of **DAAVSKK** Found in Sumo3\_with\_AS\_21Nov2010\_1.RAW  
 -IPI00303999- Tax\_Id=9606 Gene\_Symbol=PML Isoform PML-2 of Probable  
 transcription factor PML

Match to Query 1152: 1174.594872 from(392.538900,3+)



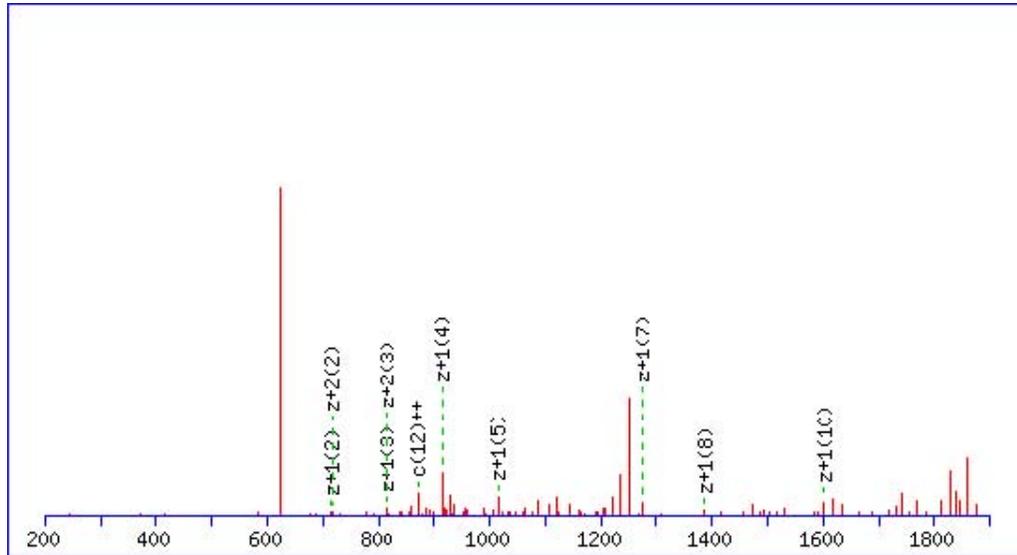
Monoisotopic mass of neutral peptide Mr(calc): 1174.59 Variable modifications: K6 :  
 Sumo 3 (K) Ions Score: 12 Expect: 44 Matches (**Bold Red**): 13/36 fragment ions using  
 40 most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	133.06	67.03	D					7
2	<b>204.10</b>	102.55	A	<b>1044.56</b>	<b>522.78</b>	1045.56	<b>523.29</b>	6
3	<b>275.13</b>	138.07	A	<b>973.52</b>	487.26	<b>974.53</b>	487.77	5
4	<b>374.20</b>	187.61	V	<b>902.48</b>	451.74	903.49	452.25	4
5	<b>461.24</b>	231.12	S	<b>803.41</b>	402.21	804.42	402.71	3
6	<b>1046.52</b>	523.76	K	<b>716.38</b>	358.69	717.39	359.20	2
7			K	131.09	66.05	132.10	66.55	1

## Peptide No.10

MS/MS Fragmentation of **STANVLEETTVKK** Found in Sumo3\_wo\_AS\_21Nov2010\_2.RAW -**IPI00290652**- Tax\_Id=9606 Gene\_Symbol=RSF1 remodeling and spacing factor 1

Match to Query 4296: 1875.952512 from(626.324780,3+)



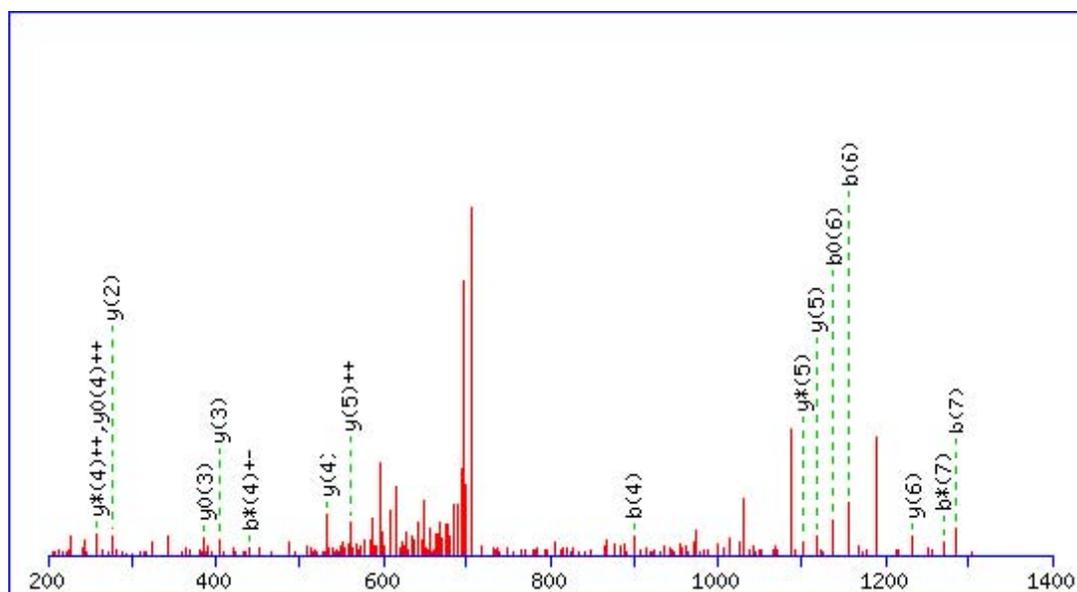
Monoisotopic mass of neutral peptide Mr(calc): 1875.95 Variable modifications: K12 : Sumo 3 (K) Ions Score: 12 Expect: 47 Matches (**Bold Red**): 10/72 fragment ions using 42 most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	105.07	53.04	S					13
2	206.11	103.56	T	1773.91	887.46	1774.92	887.96	12
3	277.15	139.08	A	1672.86	836.93	1673.87	837.44	11
4	391.19	196.10	N	<b>1601.83</b>	801.42	1602.83	801.92	10
5	490.26	245.63	V	1487.78	744.39	1488.79	744.90	9
6	603.35	302.18	L	<b>1388.71</b>	694.86	1389.72	695.36	8
7	732.39	366.70	E	<b>1275.63</b>	638.32	1276.64	638.82	7
8	861.43	431.22	E	1146.59	573.80	1147.60	574.30	6
9	962.48	481.74	T	<b>1017.54</b>	509.28	1018.55	509.78	5
10	1063.53	532.27	T	<b>916.50</b>	458.75	917.51	459.26	4
11	1162.60	581.80	V	<b>815.45</b>	408.23	<b>816.46</b>	408.73	3
12	1747.88	<b>874.44</b>	K	<b>716.38</b>	358.69	<b>717.39</b>	359.20	2
13			K	131.09	66.05	132.10	66.55	1

## Peptide No.11

MS/MS Fragmentation of **TVIKKEEK** Found in  
Sumo3\_with\_AS\_21Nov2010\_3.RAW -IPI00005648- Tax\_Id=9606  
Gene\_Symbol=SAFB2 Scaffold attachment factor B2

Match to Query 5631: 1430.774628 from(716.394590,2+)



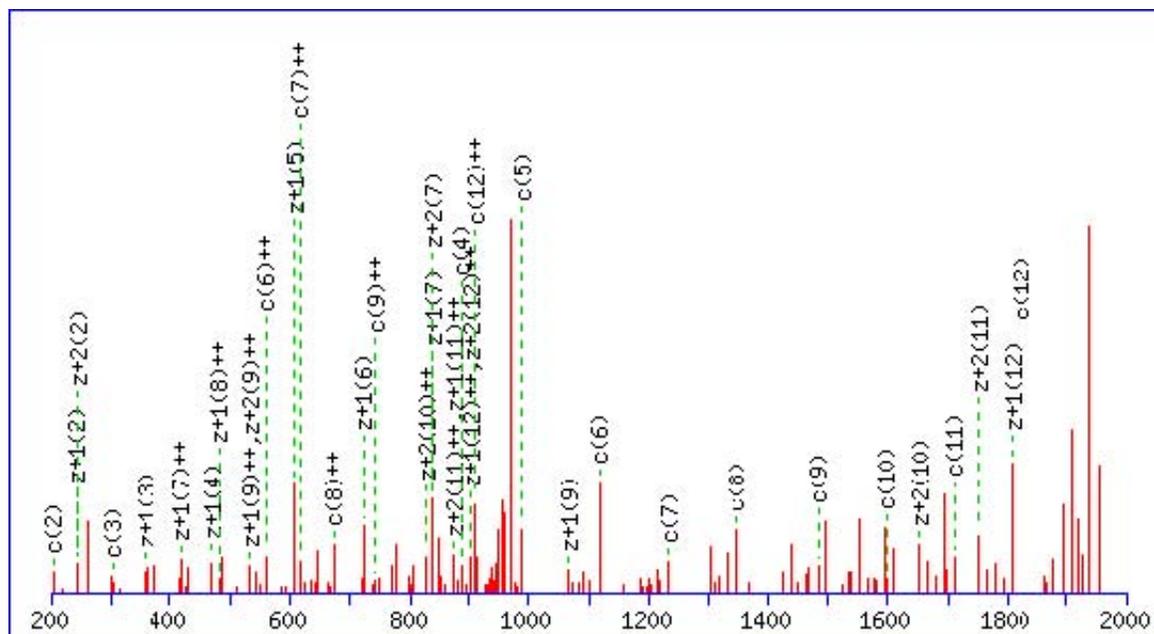
Monoisotopic mass of neutral peptide Mr(calc): 1430.77 Variable modifications: K4 :  
Sumo 3 (K) Ions Score: 13 Expect: 21 Matches (**Bold Red**): 17/76 fragment ions using 44  
most intense peaks

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	102.05	51.53			84.04	42.53	T							8
2	201.12	101.07			183.11	92.06	V	1330.73	665.87	1313.71	657.36	1312.72	656.86	7
3	314.21	157.61			296.20	148.60	I	<b>1231.66</b>	615.34	1214.64	607.82	1213.65	607.33	6
4	<b>859.49</b>	450.25	882.47	<b>441.74</b>	881.48	441.25	K	<b>1118.58</b>	<b>559.79</b>	<b>1101.55</b>	551.28	1100.57	550.79	5
5	1027.59	514.30	1010.56	505.79	1009.58	505.29	K	<b>533.29</b>	267.15	516.27	<b>258.64</b>	515.28	<b>258.14</b>	4
6	<b>1156.63</b>	578.82	1139.61	570.31	<b>1138.62</b>	569.81	E	<b>405.20</b>	203.10	388.17	194.59	<b>387.19</b>	194.10	3
7	<b>1285.67</b>	643.34	<b>1268.65</b>	634.83	1267.66	634.34	E	<b>276.16</b>	133.58	259.13	130.07	<b>258.14</b>	129.58	2
8							K	147.11	74.06	130.05	65.55			1

## Peptide No.12

MS/MS Fragmentation of **EGVKTENDHINLK** Found in  
Sumo3\_wo\_AS\_21Nov2010\_4.RAW -IPI00299147- Tax\_Id=9606  
Gene\_Symbol=SUMO3 Small ubiquitin-related modifier 3

Match to Query 4569: 1952.954616 from(489.245930,4+)



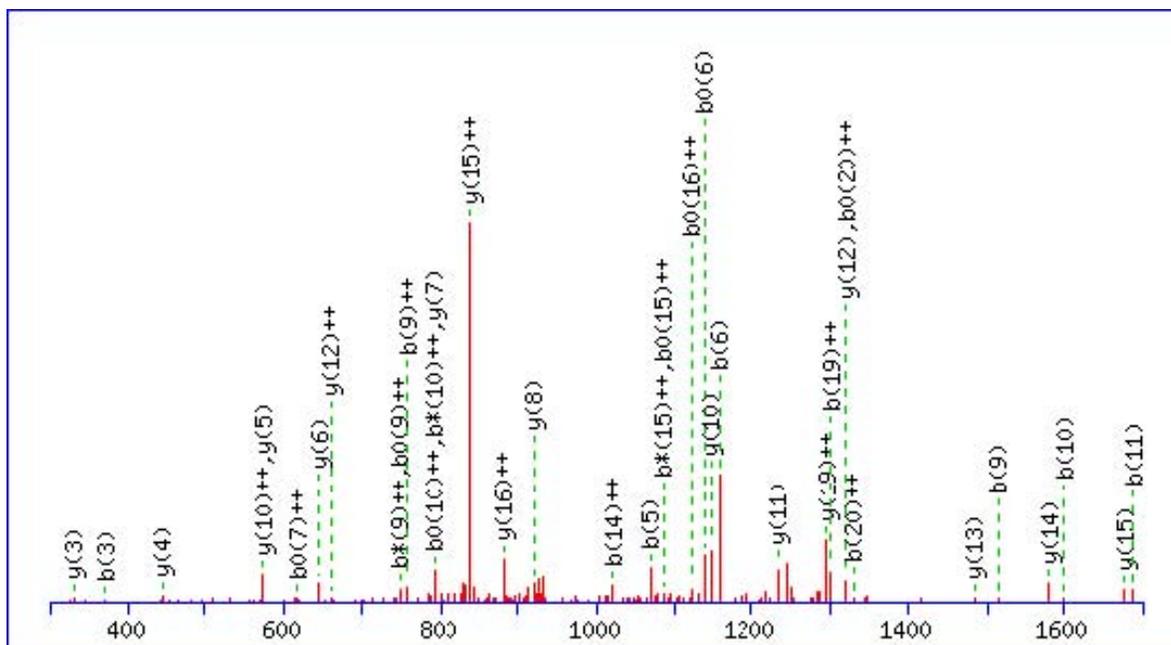
Monoisotopic mass of neutral peptide Mr(calc): 1952.96 Variable modifications: K4 :  
Sumo 3 (K) Ions Score: 32 Expect: 0.39 Matches (**Bold Red**): 37/72 fragment ions using  
112 most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	147.08	74.04	E					13
2	204.10	102.55	G	1808.90	904.95	1809.91	905.46	12
3	303.17	152.09	V	1751.88	876.44	1752.89	876.95	11
4	888.45	444.73	K	1652.81	826.91	1653.82	827.41	10
5	989.50	495.25	T	1067.52	534.27	1068.53	534.77	9
6	1118.54	559.78	E	966.48	483.74	967.48	484.25	8
7	1232.59	616.80	N	837.43	419.22	838.44	419.72	7
8	1347.61	674.31	D	723.39	362.20	724.40	362.70	6
9	1484.67	742.84	H	608.36	304.69	609.37	305.19	5
10	1597.76	799.38	I	471.31	236.16	472.31	236.66	4
11	1711.80	856.40	N	358.22	179.61	359.23	180.12	3
12	1824.88	912.95	L	244.18	122.59	245.19	123.10	2
13			K	131.09	66.05	132.10	66.55	1

### Peptide No.13

MS/MS Fragmentation of **LQEKLSPPYSSPQEFAQDVGR** Found in  
 Sumo3\_with\_AS\_21Nov2010\_5.RAW -IPI00438229- Tax\_Id=9606  
 Gene\_Symbol=TRIM28 Isoform 1 of Transcription intermediary factor 1-beta

Match to Query 12153: 2832.370932 from(945.130920,3+)



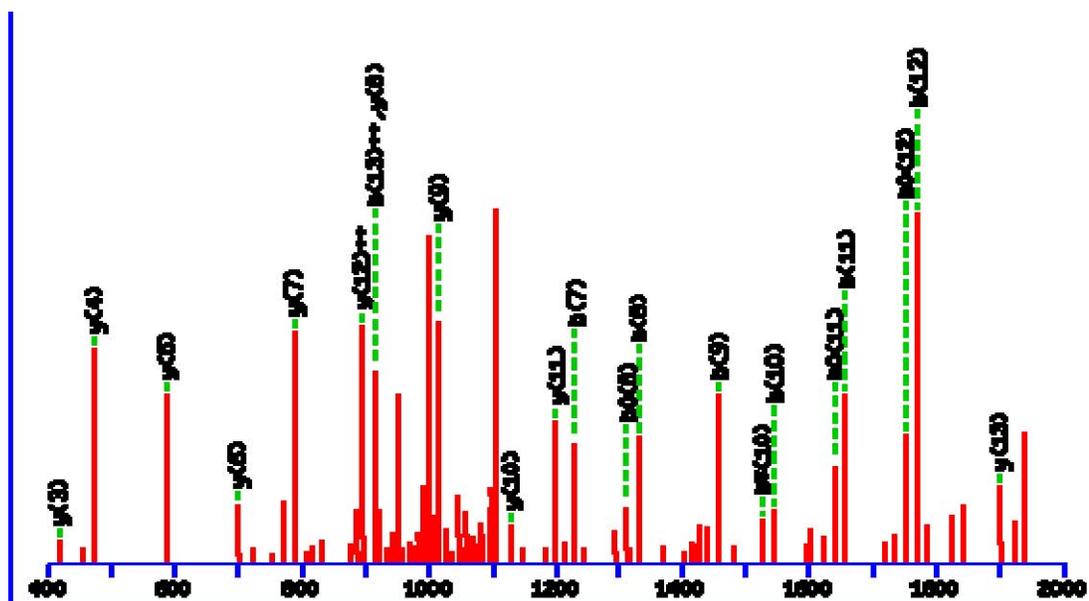
Monoisotopic mass of neutral peptide Mr(calc): 2832.37 Variable modifications: K4 :  
 Sumo 3 (K) Ions Score: 75 Expect: 4e-05 Matches (**Bold Red**): 37/228 fragment ions using  
 42 most intense peaks

#	b	b <sup>++</sup>	b <sup>*</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>*</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	114.09	57.55					L							21
2	242.15	121.58	225.12	113.07			Q	2720.29	1360.65	2703.26	1352.14	2702.28	1351.64	20
3	371.19	186.10	354.17	177.59	353.18	177.09	E	2592.23	1296.62	2575.21	1288.11	2574.22	1287.61	19
4	956.48	478.74	939.45	470.23	938.47	469.74	K	2463.19	1232.10	2446.16	1223.59	2445.18	1223.09	18
5	1069.56	535.29	1052.54	526.77	1051.55	526.28	L	1877.90	939.46	1860.88	930.94	1859.89	930.45	17
6	1156.60	578.80	1139.57	570.29	1138.59	569.80	S	1764.82	882.91	1747.79	874.40	1746.81	873.91	16
7	1253.65	627.33	1236.62	618.81	1235.64	618.32	P	1677.79	839.40	1660.76	830.88	1659.78	830.39	15
8	1350.70	675.85	1333.67	667.34	1332.69	666.85	P	1580.73	790.87	1563.71	782.36	1562.72	781.87	14
9	1513.76	757.39	1496.74	748.87	1495.75	748.38	Y	1483.68	742.34	1466.65	733.83	1465.67	733.34	13
10	1600.80	800.90	1583.77	792.39	1582.79	791.90	S	1320.62	660.81	1303.59	652.30	1302.61	651.81	12
11	1687.83	844.42	1670.80	835.90	1669.82	835.41	S	1233.59	617.30	1216.56	608.78	1215.58	608.29	11
12	1784.88	892.94	1767.85	884.43	1766.87	883.94	P	1146.55	573.78	1129.53	565.27	1128.54	564.78	10
13	1912.94	956.97	1895.91	948.46	1894.93	947.97	Q	1049.50	525.25	1032.47	516.74	1031.49	516.25	9
14	2041.98	1021.49	2024.96	1012.98	2023.97	1012.49	E	921.44	461.22	904.42	452.71	903.43	452.22	8
15	2189.05	1095.03	2172.02	1086.52	2171.04	1086.02	F	792.40	396.70	775.37	388.19	774.39	387.70	7
16	2260.09	1130.55	2243.06	1122.03	2242.08	1121.54	A	645.33	323.17	628.30	314.66	627.32	314.16	6
17	2388.15	1194.58	2371.12	1186.06	2370.14	1185.57	Q	574.29	287.65	557.27	279.14	556.28	278.65	5
18	2503.17	1252.09	2486.15	1243.58	2485.16	1243.09	D	446.24	223.62	429.21	215.11	428.23	214.62	4
19	2602.24	1301.62	2585.22	1293.11	2584.23	1292.62	V	331.21	166.11	314.18	157.59			3
20	2659.26	1330.14	2642.24	1321.62	2641.25	1321.13	G	232.14	116.57	215.11	108.06			2
21							R	175.12	88.06	158.09	79.55			1

## Peptide No.14

MS/MS Fragmentation of **LTEDKADVQSIIGLQR** Found in  
Sumo3\_wo\_AS\_21Nov2010\_4.RAW -**IPI00438229**- Tax\_Id=9606  
Gene\_Symbol=TRIM28 Isoform 1 of

Transcription intermediary factor 1-beta  
Match to Query 9541: 2242.160648 from(1122.087600,2+) Transcription intermediary  
factor 1-beta Match to Query 9541: 2242.160648 from(1122.087600,2+)



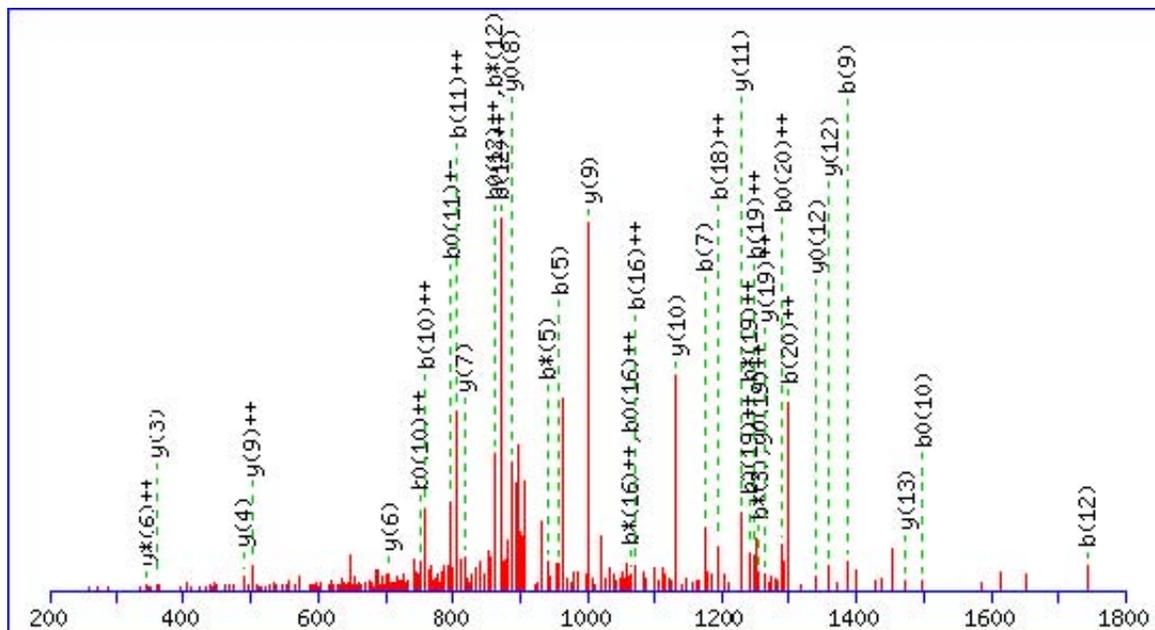
Monoisotopic mass of neutral peptide Mr(calc): 2242.16 Variable modifications: K5 :  
Sumo 3 (K) Ions Score: 70 Expect: 7.5e-05 Matches (**Bold Red**): 22/158 fragment ions  
using 33 most intense peaks

#	b	b <sup>++</sup>	b <sup>*</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>*</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	114.09	57.55					L							16
2	215.14	108.07			197.13	99.07	T	2130.08	1065.54	2113.05	1057.03	2112.07	1056.54	15
3	344.18	172.59			326.17	163.59	E	2029.03	1015.02	2012.00	1006.51	2011.02	1006.01	14
4	459.21	230.11			441.20	221.10	D	1899.99	950.50	1882.96	941.98	1881.98	941.49	13
5	1044.50	522.75	1027.47	514.24	1026.49	513.75	K	1784.96	892.98	1767.93	884.47	1766.95	883.98	12
6	1115.53	558.27	1098.51	549.76	1097.52	549.26	A	1199.67	600.34	1182.65	591.83	1181.66	591.34	11
7	1230.56	615.78	1213.53	607.27	1212.55	606.78	D	1128.64	564.82	1111.61	556.31	1110.63	555.82	10
8	1329.63	665.32	1312.60	656.80	1311.62	656.31	V	1013.61	507.31	996.58	498.80	995.60	498.30	9
9	1457.69	729.35	1440.66	720.83	1439.68	720.34	Q	914.54	457.77	897.52	449.26	896.53	448.77	8
10	1544.72	772.86	1527.69	764.35	1526.71	763.86	S	786.48	393.75	769.46	385.23	768.47	384.74	7
11	1657.80	829.41	1640.78	820.89	1639.79	820.40	I	699.45	350.23	682.42	341.72			6
12	1770.89	885.95	1753.86	877.43	1752.88	876.94	I	586.37	293.69	569.34	285.17			5
13	1827.91	914.46	1810.88	905.94	1809.90	905.45	G	473.28	237.15	456.26	228.63			4
14	1940.99	971.00	1923.97	962.49	1922.98	961.99	L	416.26	208.63	399.24	200.12			3
15	2069.05	1035.03	2052.02	1026.52	2051.04	1026.02	Q	303.18	152.09	286.15	143.58			2
16							R	175.12	88.06	158.09	79.55			1

## Peptide No.15

MS/MS Fragmentation of **TLTGKTITLEVEPSDTIENVK** Found in Sumo3\_with\_AS\_21Nov2010\_4.RAW -IPI00456429- Tax\_Id=9606 Gene\_Symbol=UBA52;UBB;RPS27A; UBC ubiquitin and ribosomal protein L40 precursor

Match to Query 12073: 2744.410632 from(915.810820,3+)



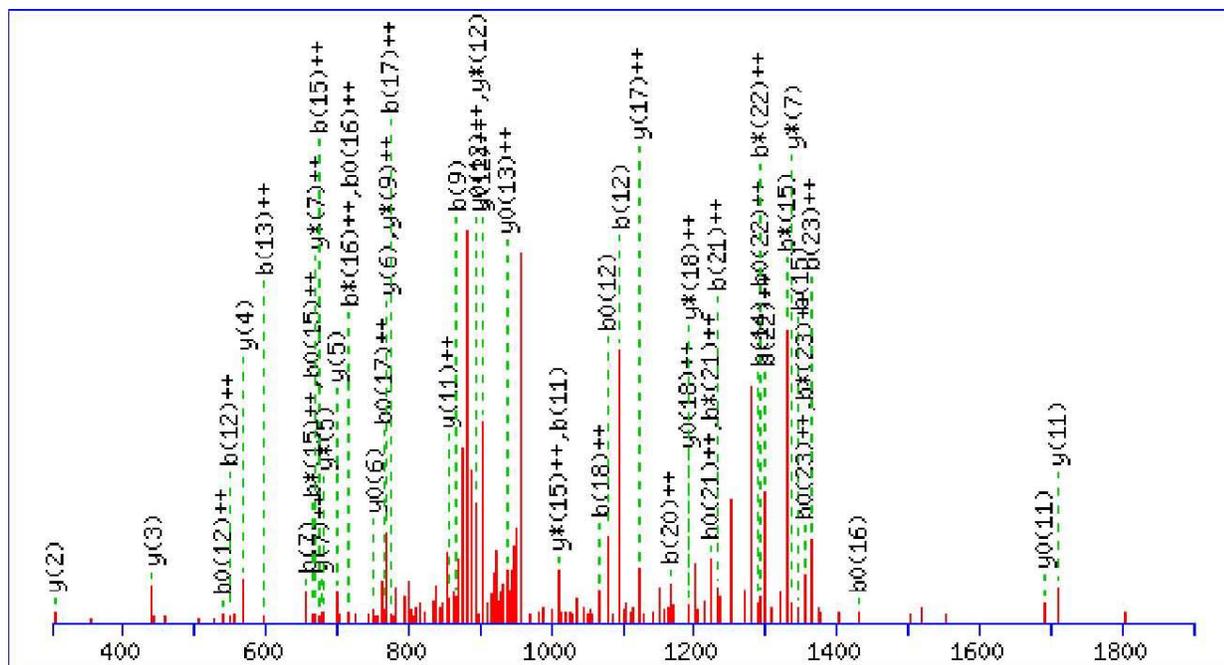
Monoisotopic mass of neutral peptide Mr(calc): 2744.41 Variable modifications: K5 : Sumo 3 (K) Ions Score: 28 Expect: 1.2 Matches (**Bold Red**): 38/226 fragment ions using 83 most intense peaks

#	b	b <sup>++</sup>	b <sup>+</sup>	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y <sup>+</sup>	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	102.05	51.53			84.04	42.53	T							21
2	215.14	108.07			197.13	99.07	L	2644.37	1322.69	2627.34	1314.17	2626.36	1313.68	20
3	316.19	158.60			298.18	149.59	T	2531.28	1266.15	2514.26	1257.63	2513.27	1257.14	19
4	373.21	187.11			355.20	178.10	G	2430.24	1215.62	2413.21	1207.11	2412.23	1206.62	18
5	958.50	479.75	941.47	471.24	940.48	470.75	K	2373.21	1187.11	2356.19	1178.60	2355.20	1178.11	17
6	1059.54	530.28	1042.52	521.76	1041.53	521.27	T	1787.93	894.47	1770.90	885.95	1769.92	885.46	16
7	1172.63	586.82	1155.60	578.30	1154.62	577.81	I	1686.88	843.94	1669.85	835.43	1668.87	834.94	15
8	1273.67	637.34	1256.65	628.83	1255.66	628.34	T	1573.80	787.40	1556.77	778.89	1555.78	778.40	14
9	1386.76	693.88	1369.73	685.37	1368.75	684.88	L	1472.75	736.88	1455.72	728.36	1454.74	727.87	13
10	1515.80	758.40	1498.77	749.89	1497.79	749.40	E	1359.66	680.34	1342.64	671.82	1341.65	671.33	12
11	1614.87	807.94	1597.84	799.43	1596.86	798.93	V	1230.62	615.81	1213.59	607.30	1212.61	606.81	11
12	1743.91	872.46	1726.89	863.95	1725.90	863.45	E	1131.55	566.28	1114.53	557.77	1113.54	557.27	10
13	1840.97	920.99	1823.94	912.47	1822.95	911.98	P	1002.51	501.76	985.48	493.25	984.50	492.75	9
14	1928.00	964.50	1910.97	955.99	1909.99	955.50	S	905.46	453.23	888.43	444.72	887.45	444.23	8
15	2043.02	1022.02	2026.00	1013.50	2025.01	1013.01	D	818.43	409.72	801.40	401.20	800.41	400.71	7
16	2144.07	1072.54	2127.05	1064.03	2126.06	1063.53	T	703.40	352.20	686.37	343.69	685.39	343.20	6
17	2257.16	1129.08	2240.13	1120.57	2239.15	1120.08	I	602.35	301.68	585.32	293.17	584.34	292.67	5
18	2386.20	1193.60	2369.17	1185.09	2368.19	1184.60	E	489.27	245.14	472.24	236.62	471.26	236.13	4
19	2500.24	1250.62	2483.21	1242.11	2482.23	1241.62	N	360.22	180.62	343.20	172.10			3
20	2599.31	1300.16	2582.28	1291.65	2581.30	1291.15	V	246.18	123.59	229.15	115.08			2
21							K	147.11	74.06	130.09	65.55			1

## Peptide No.16

MS/MS Fragmentation of **AADGGERPLAASPPGTVKAEEHQ**R Found in Sumo3\_with\_AS\_21Nov2010\_6.RAW -IPI00295502- Tax\_Id=9606 Gene\_Symbol=WIZ Isoform 1 of Protein Wiz

Match to Query 12225: 2900.416002 from(967.812610,3+)



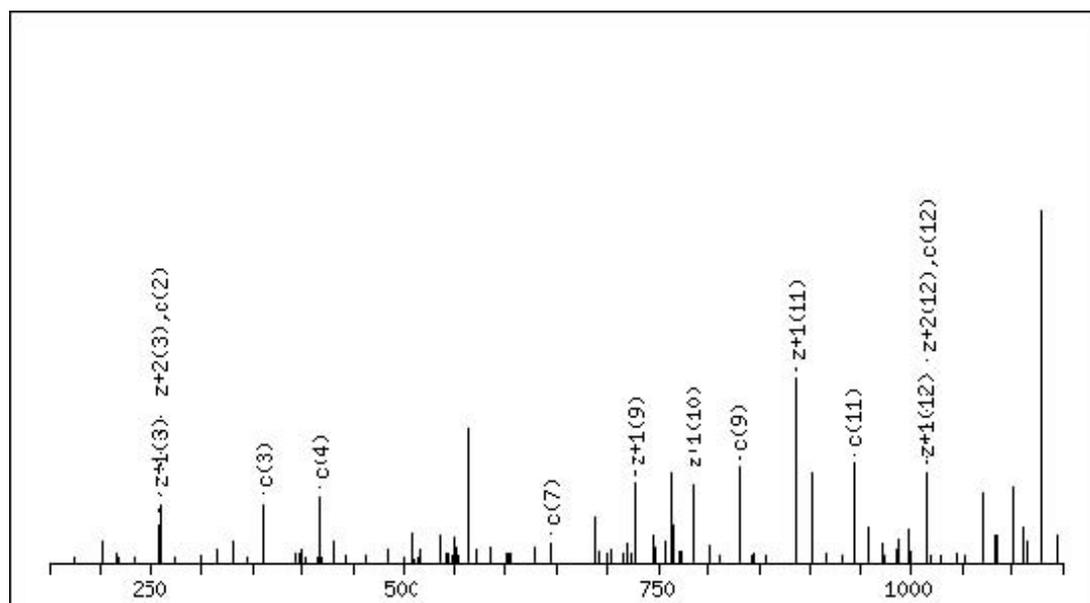
Monoisotopic mass of neutral peptide Mr(calc): 2900.41 Variable modifications: K18 : Sumo 3 (K) Ions Score: 29 Expect: 1.6 Matches (**Red**): 52/254 fragment ions using 99 most intense peaks

#	<b>b</b>	<b>b<sup>++</sup></b>	<b>b<sup>*</sup></b>	<b>b<sup>*++</sup></b>	<b>b<sup>0</sup></b>	<b>b<sup>0++</sup></b>	Seq.	<b>y</b>	<b>y<sup>++</sup></b>	<b>y<sup>*</sup></b>	<b>y<sup>*++</sup></b>	<b>y<sup>0</sup></b>	<b>y<sup>0++</sup></b>
1	72.04	36.53					A						
2	143.08	72.04					A	2830.38	1415.69	2813.36	1407.18	2812.37	1406.69
3	258.11	129.56			240.10	120.55	D	2759.35	1380.18	2742.32	1371.66	2741.33	1371.17
4	315.13	158.07			297.12	149.06	G	2644.32	1322.66	2627.29	1314.15	2626.31	1313.66
5	372.15	186.58			354.14	177.57	G	2587.30	1294.15	2570.27	1285.64	2569.29	1285.15
6	501.19	251.10			483.18	242.10	E	2530.28	1265.64	2513.25	1257.13	2512.27	1256.64
7	657.30	329.15	640.27	320.64	639.28	320.15	R	2401.23	1201.12	2384.21	1192.61	2383.22	1192.11
8	754.35	377.68	737.32	369.16	736.34	368.67	P	2245.13	1123.07	2228.11	1114.56	2227.12	1114.06
9	867.43	434.22	850.41	425.71	849.42	425.21	L	2148.08	1074.54	2131.05	1066.03	2130.07	1065.54
10	938.47	469.74	921.44	461.22	920.46	460.73	A	2035.00	1018.00	2017.97	1009.49	2016.98	1009.00
11	1009.51	505.26	992.48	496.74	991.50	496.25	A	1963.96	982.48	1946.93	973.97	1945.95	973.48
12	1096.54	548.77	1079.51	540.26	1078.53	539.77	S	1892.92	946.96	1875.89	938.45	1874.91	937.96
13	1193.59	597.30	1176.56	588.79	1175.58	588.29	P	1805.89	903.45	1788.86	894.93	1787.88	894.44
14	1290.64	645.83	1273.62	637.31	1272.63	636.82	P	1708.84	854.92	1691.81	846.41	1690.83	845.92
15	1347.67	674.34	1330.64	665.82	1329.65	665.33	G	1611.78	806.40	1594.76	797.88	1593.77	797.39
16	1448.71	724.86	1431.69	716.35	1430.70	715.85	T	1554.76	777.88	1537.74	769.37	1536.75	768.88
17	1547.78	774.39	1530.75	765.88	1529.77	765.39	V	1453.71	727.36	1436.69	718.85	1435.70	718.36
18	2133.07	1067.04	2116.04	1058.52	2115.06	1058.03	K	1354.65	677.83	1337.62	669.31	1336.64	668.82
19	2204.11	1102.56	2187.08	1094.04	2186.09	1093.55	A	769.36	385.18	752.33	376.67	751.35	376.18
20	2333.15	1167.08	2316.12	1158.56	2315.14	1158.07	E	698.32	349.66	681.30	341.15	680.31	340.66
21	2462.19	1231.60	2445.16	1223.09	2444.18	1222.59	E	569.28	285.14	552.25	276.63	551.27	276.14
22	2599.25	1300.13	2582.22	1291.62	2581.24	1291.12	H	440.24	220.62	423.21	212.11		
23	2727.31	1364.16	2710.28	1355.64	2709.30	1355.15	Q	303.18	152.09	286.15	143.58		
24							R	175.12	88.06	158.09	79.55		

**Peptide No.17**

MS/MS Fragmentation of **NQTGGGLGKGGAK** Found in  
Sumo3\_with\_AS\_21Nov2010\_3.RAW -IPI00453473- Tax\_Id=9606  
Gene\_Symbol=HIST1H4H;HIST2H4A;HIST4H4;HIST1H4F;HIST1H4D;HIST1H4K;HIS  
T1H4C;HIST1H4J;HIST1H4A;HIST1H4I;HIST1H4B;HIST2H4B;HIST1H4E;HIST1H4L  
Histone H4

Match to Query 1020: 1143.601482 from(382.207770,3+)



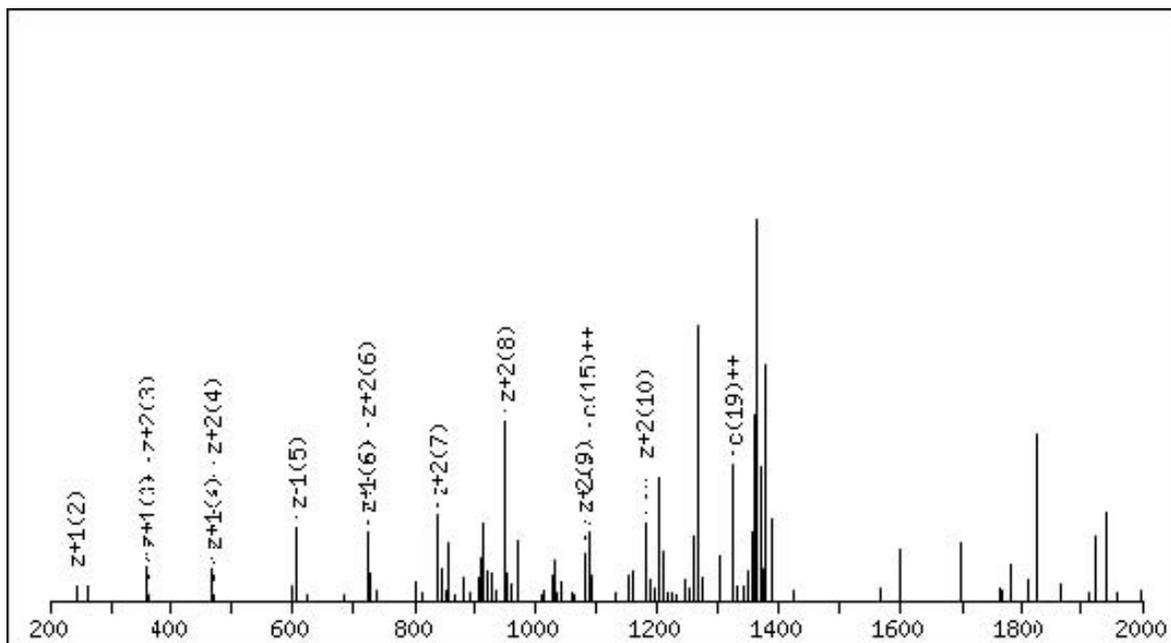
Monoisotopic mass of neutral peptide Mr(calc): 1143.60 Ions Score: 53 Expect: 0.001  
Matches (**Bold Red**): 14/72 fragment ions using 21 most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	132.08	66.54	N					13
2	260.14	130.57	Q	1014.55	507.78	1015.55	508.28	12
3	361.18	181.10	T	886.49	443.75	887.49	444.25	11
4	418.20	209.61	C	785.44	393.22	786.45	393.73	10
5	475.23	238.12	C	728.42	364.71	729.43	365.22	9
6	532.25	266.63	C	671.40	336.20	672.40	336.71	8
7	645.33	323.17	L	614.37	307.69	615.38	308.19	7
8	702.35	351.68	C	501.29	251.15	502.30	251.65	6
9	830.45	415.73	K	444.27	222.64	445.28	223.14	5
10	887.47	444.24	C	316.17	158.59	317.18	159.09	4
11	944.49	472.75	C	259.15	130.08	260.16	130.58	3
12	1015.53	508.27	A	202.13	101.57	203.14	102.07	2
13			K	131.09	66.05	132.10	66.55	1

**Peptide No.18**

MS/MS Fragmentation of **ANEKPTEEVKTENNNHINLK** Found in Sumo3\_wo\_AS\_21Nov2010\_4.RAW -**IPI00434968**- Small ubiquitin-related modifier 4

Match to Query 5418: 2778.362616 from(695.597930,4+)



Monoisotopic mass of neutral peptide Mr(calc): 2778.35 Variable modifications: K10 : Sumo 3 (K) Ions Score: 22 Expect: 7.5 Matches (**Bold Red**): 14/114 fragment ions using 37 most intense peaks

#	c	c <sup>++</sup>	Seq.	z+1	z+1 <sup>++</sup>	z+2	z+2 <sup>++</sup>	#
1	89.07	45.04	A					20
2	203.11	102.06	N	2692.30	1346.66	2693.31	1347.16	19
3	332.16	166.58	E	2578.26	1289.63	2579.27	1290.14	18
4	460.25	230.63	K	2449.22	1225.11	2450.23	1225.62	17
5	557.30	279.16	P	2321.12	1161.07	2322.13	1161.57	16
6	658.35	329.68	T	2224.07	1112.54	2225.08	1113.04	15
7	787.39	394.20	E	2123.02	1062.02	2124.03	1062.52	14
8	916.44	458.72	E	1993.98	997.49	1994.99	998.00	13
9	1015.51	508.26	V	1864.94	932.97	1865.95	933.48	12
10	1600.79	800.90	K	1765.87	883.44	1766.88	883.94	11
11	1701.84	851.42	T	1180.58	590.80	1181.59	591.30	10
12	1830.88	915.95	E	1079.54	540.27	1080.54	540.78	9
13	1944.93	972.97	N	950.49	475.75	951.50	476.25	8
14	2058.97	1029.99	N	836.45	418.73	837.46	419.23	7
15	2173.01	1087.01	N	722.41	361.71	723.41	362.21	6
16	2310.07	1155.54	H	608.36	304.69	609.37	305.19	5
17	2423.15	1212.08	I	471.31	236.16	472.31	236.66	4
18	2537.20	1269.10	N	358.22	179.61	359.23	180.12	3
19	2650.28	1325.64	L	244.18	122.59	245.19	123.10	2
20			K	131.09	66.05	132.10	66.55	1

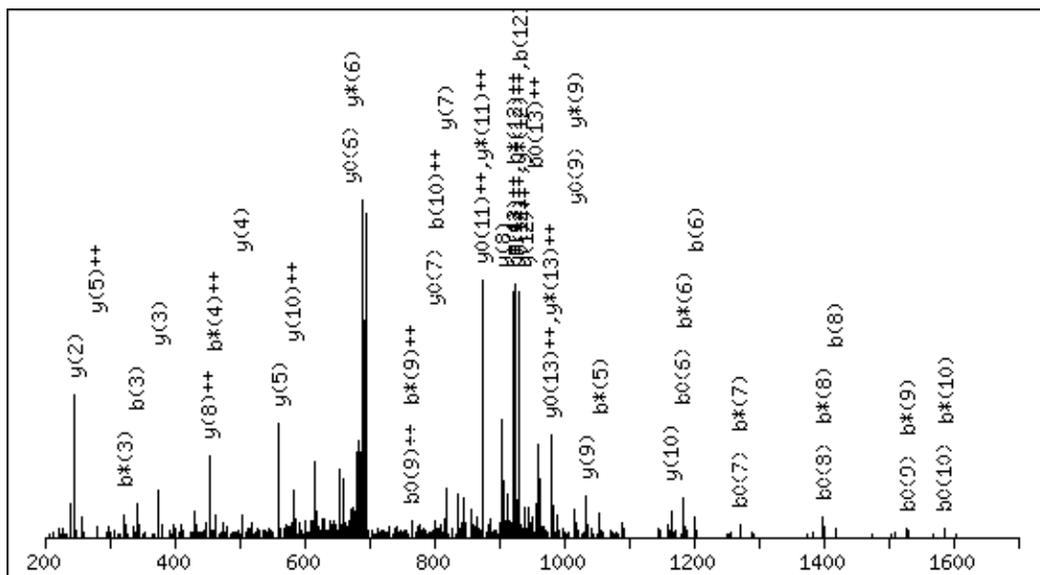
**Annex III: Supplementary Figure 2.11 for Chapter 2**  
**(A Novel Proteomics Approach to Identify SUMOylated Proteins and**  
**their Modification Sites in Human Cells)**

**Supplementary Figure 2.11:** Tandem Mass Spectra of All Identified SUMO-1 Peptides Following Database Search and Manual Validation

### Peptide No.1

#### MS/MS Fragmentation of **KVIKMESEEGKEAR**

Found in **IPI:IPI00022348**, Tax\_Id=9606 Gene\_Symbol=PML Isoform PML-1 of Probable transcription factor PML



Monoisotopic mass of neutral peptide Mr(calc): 2105.04

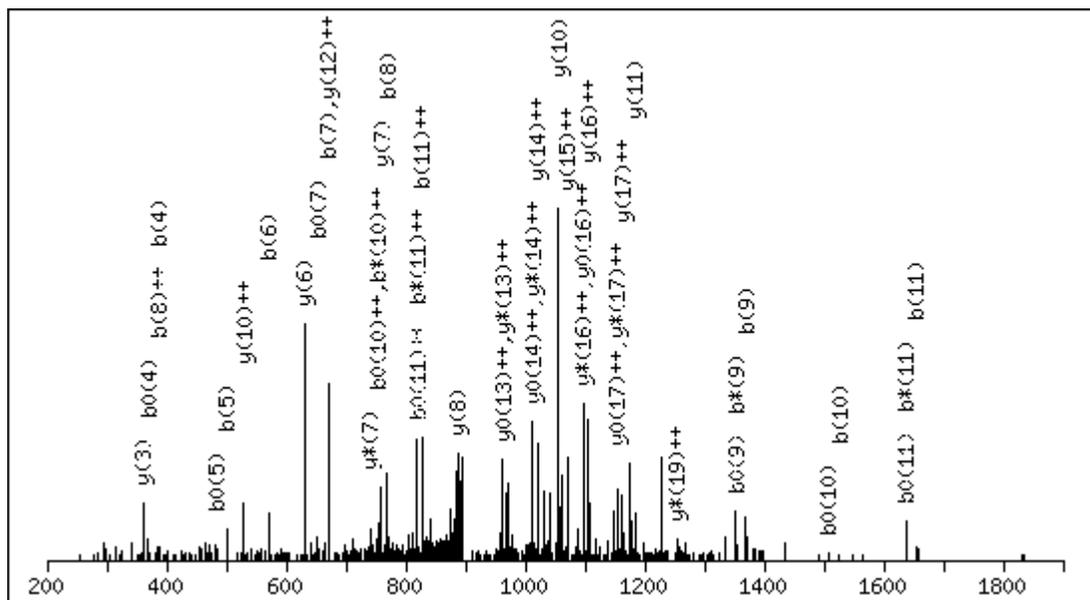
Variable modifications:

K4 : SUMO EQTGG (K) (K)

Ions Score: 30 Expect: 0.81

Matches (**Bold Red**): 47/142 fragment ions using 105 most intense peaks

#	b	b <sup>++</sup>	b*	b <sup>*++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>*++</sup>	y <sup>0</sup>	y <sup>0++</sup>
1	129.1	65.05	112.08	56.54			K						
2	228.17	114.59	211.14	106.08			V	1977.95	989.48	1960.93	<b>980.97</b>	1959.94	<b>980.48</b>
3	<b>341.25</b>	171.13	<b>324.23</b>	162.62			I	1878.89	<b>939.95</b>	1861.86	<b>931.43</b>	1860.88	<b>930.94</b>
4	941.54	471.27	<b>924.51</b>	<b>462.76</b>			K	1765.8	883.4	1748.78	<b>874.89</b>	1747.79	<b>874.4</b>
5	1072.58	536.79	<b>1055.56</b>	528.28			M	<b>1165.52</b>	<b>583.26</b>	1148.49	574.75	1147.5	574.26
6	<b>1201.62</b>	601.32	<b>1184.6</b>	592.8	<b>1183.61</b>	592.31	E	<b>1034.47</b>	517.74	<b>1017.45</b>	509.23	<b>1016.46</b>	508.74
7	1288.66	644.83	<b>1271.63</b>	636.32	<b>1270.65</b>	635.83	S	<b>905.43</b>	<b>453.22</b>	888.41	444.71	887.42	444.21
8	<b>1417.7</b>	709.35	<b>1400.67</b>	700.84	<b>1399.69</b>	700.35	E	<b>818.4</b>	409.7	801.37	401.19	<b>800.39</b>	400.7
9	1546.74	773.87	<b>1529.72</b>	<b>765.36</b>	<b>1528.73</b>	<b>764.87</b>	E	689.36	345.18	<b>672.33</b>	336.67	<b>671.35</b>	336.18
10	1603.76	<b>802.39</b>	<b>1586.74</b>	793.87	<b>1585.75</b>	793.38	G	<b>560.32</b>	<b>280.66</b>	543.29	272.15	542.3	271.66
11	1731.86	866.43	1714.83	857.92	1713.85	857.43	K	<b>503.29</b>	252.15	486.27	243.64	485.28	243.15
12	1860.9	<b>930.95</b>	1843.87	<b>922.44</b>	1842.89	<b>921.95</b>	E	<b>375.2</b>	188.1	358.17	179.59	357.19	179.1
13	1931.94	966.47	1914.91	957.96	1913.93	<b>957.47</b>	A	<b>246.16</b>	123.58	229.13	115.07		
14							R	175.12	88.06	158.09	79.55		

**Peptide No. 2**MS/MS Fragmentation of **ADSLLA**VVKREPAEQGDGERFound in **IPI:IPI00005648**, Tax\_Id=9606 Gene\_Symbol=SAFB2 Scaffold attachment factor B2

Monoisotopic mass of neutral peptide Mr(calc): 2708.34

Variable modifications:

K9 : SUMO EQTGG (K) (K)

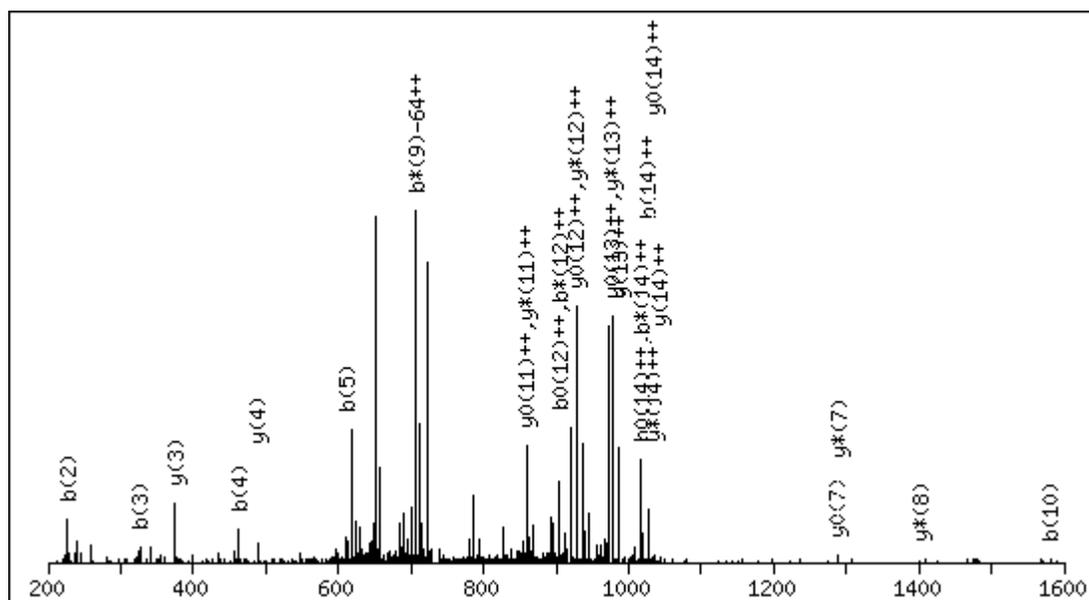
Ions Score: 30 Expect: 1

Matches (**Bold Red**): 44/220 fragment ions using 75 most intense peaks

#		$b^{++}$	$b^*$	$b^{*++}$	$b^0$	$b^{0++}$	Seq.	y	$y^{++}$	$y^{*++}$	$y^0$	$y^{0++}$
1	72.04	36.53					A					
2	187.07	94.04			169.06	85.03	D	2638.31	1319.66	1311.14	2620.3	1310.65
3	274.1	137.56			256.09	128.55	S	2523.28	1262.14	1253.63	2505.27	1253.14
4	387.19	194.1			369.18	185.09	L	2436.25	1218.63	1210.11	2418.24	1209.62
5	500.27	250.64			482.26	241.63	L	2323.16	1162.09	1153.57	2305.15	1153.08
6	571.31	286.16			553.3	277.15	A	2210.08	1105.54	1097.03	2192.07	1096.54
7	670.38	335.69			652.37	326.69	V	2139.04	1070.02	1061.51	2121.03	1061.02
8	769.45	385.23			751.43	376.22	V	2039.97	1020.49	1011.98	2021.96	1011.49
9	1369.73	685.37	1352.71	676.86	1351.72	676.36	K	1940.91	970.96	962.44	1922.9	961.95
10	1525.83	763.42	1508.81	754.91	1507.82	754.41	R	1340.62	670.81	662.3	1322.61	661.81
11	1654.88	827.94	1637.85	819.43	1636.87	818.94	E	1184.52	592.76	584.25	1166.51	583.76
12	1751.93	876.47	1734.9	867.95	1733.92	867.46	P	1055.48	528.24	519.73	1037.46	519.24
13	1822.97	911.99	1805.94	903.47	1804.96	902.98	A	958.42	479.71	471.2	940.41	470.71
14	1952.01	976.51	1934.98	967.99	1934	967.5	E	887.39	444.2	435.68	869.37	435.19
15	2080.07	1040.54	2063.04	1032.02	2062.06	1031.53	Q	758.34	379.68	371.16	740.33	370.67
16	2177.12	1089.06	2160.09	1080.55	2159.11	1080.06	P	630.28	315.65	307.13	612.27	306.64
17	2234.14	1117.57	2217.11	1109.06	2216.13	1108.57	G	533.23	267.12	258.61	515.22	258.11
18	2349.17	1175.09	2332.14	1166.57	2331.16	1166.08	D	476.21	238.61	230.1	458.2	229.6
19	2406.19	1203.6	2389.16	1195.09	2388.18	1194.59	G	361.18	181.1	172.58	343.17	172.09
20	2535.23	1268.12	2518.21	1259.61	2517.22	1259.11	E	304.16	152.58	144.07	286.15	143.58
21							R	175.12	88.06	79.55		

### Peptide No. 3

MS/MS Fragmentation of **LLVHMGLLKSEDKVK**  
 Found in **IPI:IPI00294879**, Tax\_Id=9606 Gene\_Symbol=RANGAP1 Ran GTPase-activating protein 1



Monoisotopic mass of neutral peptide Mr(calc): 2197.18

Variable modifications: M5 : Oxidation (M), with neutral losses 0.00(shown in table), 64.00 K9 : SUMO EQTGG (K) (K) Ions Score: 17 Expect: 7.6 Matches (**Bold Red**): 26/194 fragment ions using 45 most intense peaks

#	b	b <sup>++</sup>	b*	b <sup>*++</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>*++</sup>	y <sup>0</sup>
1	114.09	57.55				L					
2	<b>227.18</b>	114.09				L	2085.1	<b>1043.05</b>	2068.07	<b>1034.54</b>	2067.09
3	<b>326.24</b>	163.63				V	1972.02	<b>986.51</b>	1954.99	<b>978</b>	1954.01
4	<b>463.3</b>	232.16				H	1872.95	936.98	1855.92	<b>928.46</b>	1854.94
5	<b>610.34</b>	305.67				M	1735.89	868.45	1718.86	<b>859.94</b>	1717.88
6	667.36	334.18				G	1588.85	794.93	1571.83	786.42	1570.84
7	780.44	390.73				L	1531.83	766.42	1514.81	757.91	1513.82
8	893.53	447.27				L	1418.75	709.88	<b>1401.72</b>	701.36	1400.74
9	1493.81	747.41	1476.79	738.9		K	1305.66	653.34	<b>1288.64</b>	644.82	<b>1287.65</b>
10	<b>1580.85</b>	790.93	1563.82	782.41	781.92	S	705.38	353.19	688.35	344.68	687.37
11	1709.89	855.45	1692.86	846.93	846.44	E	618.35	309.68	601.32	301.16	600.34
12	1824.92	912.96	1807.89	<b>904.45</b>	<b>903.96</b>	D	<b>489.3</b>	245.16	472.28	236.64	471.29
13	1953.01	977.01	1935.98	968.5	968	K	<b>374.28</b>	187.64	357.25	179.13	
14	2052.08	<b>1026.54</b>	2035.05	<b>1018.03</b>	<b>1017.54</b>	V	246.18	123.59	229.15	115.08	
15						K	147.11	74.06	130.09	65.55	

**Annex IV: Supplementary Tables I-III for Chapter 2**  
**(A Novel Proteomics Approach to Identify SUMOylated Proteins and**  
**their Modification Sites in Human Cells)**

Supplementary Table I: Peptide and protein identification from NTA affinity-purified mock HEK293 and HEK293 His<sub>6</sub>-SUMO3 mutant nuclear extracts with and without As<sub>2</sub>O<sub>3</sub>. Table on CD ROM

Supplementary Table II: Peptide and protein identification from dual affinity purification (NTA and immunoprecipitation) of mock HEK293 and HEK293 His<sub>6</sub>-SUMO1 mutant nuclear extracts with and without As<sub>2</sub>O<sub>3</sub>. Table on CD ROM

Supplementary Table III : Identification of SUMOylated peptides with Mascot, ChopNSpice and SUMmOn from NTA affinity-purified HEK293 His<sub>6</sub>-SUMO3 mutant nuclear extracts with and without As<sub>2</sub>O<sub>3</sub>.

Protein	Peptide Sequence	K	MASCOT*		ChopNSpice		SUMmOn CID
			CID	ETD	CID	ETD	
Histone H3.2	KQLATKAAR	K6				17.7	
HSF4B	REKGLALLK	K3	37.8				
Lamin A/C	KLESTESR	K1		18.7		48.8	
PML	KVIKMESEEGKEAR	K4	46.31	39	59.7	60.23	
PML	KVIKMESEEGKEAR	K4 K11		15.5			
PML	DAAVSKK	K6		12.2			
PML	TDGFDEFKVR	K8	25.7				
RSF1	STANVLEETTVKK	K12					
SAFB2	TVIKKEEK	K4	12.9		45.38		
SUMO3	EGVKTENDHINLK	K4	19		29.5		
TRIM28	LQEKLSPPYSSPQEFAQDVGR	K4	74.9		81.14		
TRIM29	LTEDKADVQSIIGLQR	K5	48.3		68		
WIZ 1	AADGGERPLAASPPGTVKAEHQ	K18	28.8	20.9	32.43		
Ubiquitin	TLTGKTITLEVEPSDTIENVK	K5	28.2		32.9		
SUMO2/3	HTPLSKLMK	K6	20.2				
SUMO4	ANEKPTEEVKTENNNHINLK	K10		22			
SCIN	GKDANPQERK	K2			51		
protein LOC205717	KIGINR	K1			43		
Histone H4	GLGKGGAK	K4				52.6	
PML	TDGFDEFKVR	K8					0.987
CHD-7	EDVEKNLAPK	K5					0.989