Conjugation and Deconjugation of the Small Ubiquitin-like Modifiers SUMO-1, SUMO-2 and SUMO-3

Michael Howard Tatham

A Thesis Presented for the Degree of

Doctor of Philosophy

School of Biological and Medical Sciences

University of St. Andrews

February 2001

DECLARATION

I, Michael Howard Tatham, hereby certify that this thesis, which is approximately 40000 words in length, has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for higher degree.

Date 21/5/01 Signature of candidate

I was admitted as research student in September 1997 and as a candidate for the degree of Doctor of Philosophy in September 1997: the higher study for which this is a record was carried out in the faculty of Sciences of the University of St. Andrews between 1997 and 2000.

Date 21/5/01 Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 21/5/01 Signature of supervisor

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date 21/5/01 Signature of candidate

CONTENTS

List of Figures

List of Tables

List of Abbreviations

ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
APC	Anaphase promoting complex
APL	Acute promyelocytic leukaemia
APP-BP1	Amyloid precursor protein-binding protein 1
A. thaliana	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
bp	Base pairs
BPV	Bovine papilloma virus
BSA	Bovine serum albumen
CDK	Cyclin-dependant kinase
cDNA	Complementary DNA
CMV	Cytomegalo virus
CoA	Coenzyme A
COSY	Correlated spectroscopy
CSI	Chemical shift index
DeUb	De-ubiquitination
D-MEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonuleic acid
DTT	Dithiothrietol
DUB	Deubiquitination enzymes
E. coli	*Eschericia coli*
E1	Activating enzyme
E2	Conjugating enzyme
E3	Ligase enzyme
E6-AP	E6 activating protein
ECL	Enhanced chemiluminesence
EDTA	Ethylenediaminetetracetic acid
EGF	Epidermal growth factor
FAT	Factor acetyl-transferase
FCS	Foetal calf serum
GEF	Guanidine nucleotide exchange factor
GlcNAc	N-acetylglucosamine
GLUT	Glucose transporter protein
GPI	Glycophosphatidyl inositol
GST	Glutathione S-transferase

h	Hour(s)
HA	Haemagglutinin
HAT	Histone acetyl-transferase
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HECT	Homology to E6-AP C-terminus
HIPK	Homeodomain-interacting protein kinase
HLH	Helix-loop-helix
HNCA	Hydrogen-nitrogen carbon-alpha
HN(CO)CA	Hydrogen-nitrogen (carboxy-carbon) carbon-alpha
HPV	Human papilloma virus
HSQC	Heteronuclear-single-quantum coherence
Ig	Immunoglobulin
IκB	Inhibitor kappa B
INCL	Infantile neuronal ceroid lipofucinosis
IPTG	Isopropyl-β-D-thiogalactopyranose
IVTT	*In vitro* transcription and translation
KCl	Potassium chloride
kDa	Kilo Dalton
LB	Luria-Bertani broth
Mdm2	Mouse double minute 2
MHC	Major histocompatibility complex
min	Minute(s)
MonoUb	Mono-ubiquitination
MW	Molecular weight
NBs	PML nuclear bodies
NEDD8	Neuronal precursor cell expressed developmentally downregulated protein 8
NES	Nuclear export sequence
NF-κB	Nuclear Factor-Kappa B
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser enhancement spectroscopy
NP-40	Nonidet P-40
NPC	Nuclear pore complex
ODC	Ornithine decarboxylase
O-GlcNAc-T	N-acetylglucosaminyl-transferase
O-GlcNAcase	β-D-N-acetylglucosaminidase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pi	Phosphate

PDGF	Platelet-derived growth factor
PML	Promyelocytic leukaemia protein
PolyUb	Poly-ubiquitination
ppm	Parts per million
PPT	Palmitoyl-protein-thioesterases
PVDF	Polyvinylidene difluoride
RanGAP	Ran GTPase activating protein
RanBP	Ran binding protein
RAR	Retinoic acid receptor
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase PCR
Rub1	Related to ubiquitin
SAE1	SUMO-1 activating enzyme subunit 1
SAE2	SUMO-1 activating enzyme subunit 2
S. cerevisiae	*Saccharomyces cerevisiae*
SCF	Skp1, cullin, F-box
SDS	Sodium dodecyl sulfate
SENP	Sentrin-specific protease
SH2	src homology region
SUMO	Small ubiquitin-like modifier
SUSP	SUMO-specific protease
SV40	Simian virus type 40
TEL	E-26 transforming specific-related gene
3D	Three-dimensional
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
2D	Two-dimensional
Ub	Ubiquitin
UBA1	Ubiquitin activating enzyme
Ubc	Ubiquitin conjugating enzyme
UBL	Ubiquitin-like modifiers
UBP	Ubiquitin-specific processing proteases
UCH	Ubiquitin C-terminal hydrolase
UD	Ubiquitin domain
UDP	Ubiquitin-domain protein
ULP	Ubiquitin-like protein
Ulp	Ubiquitin-like protease
UV	Ultra-violet
VDW	Van Der Waals
VHL	Von Hippel-Lindau
Wrn	Werner's disease protein
WT	Wild-type

Abbreviations for Amino Acids and Side-chains

Alanine	ala	A	$-CH_3$
Arginine	arg	R	$-(CH_2)_3-NH-CNH-NH_2$
Asparagine	asn	N	$-CH_2-CONH_2$
Aspartic acid	asp	D	$-CH_2-COOH$
Cysteine	cys	C	$-CH_2-SH$
Glutamine	gln	Q	$-CH_2-CH_2-CONH_2$
Glutamic acid	glu	E	$-CH_2-CH_2-COOH$
Glycine	gly	G	-H
Histidine	his	H	$-CH_2$-imidazole
Isoleucine	ile	I	$-CH(CH_3)-CH_2-CH_3$
Leucine	leu	L	$-CH_2-CH(CH_3)_2$
Lysine	lys	K	$-(CH_2)_4-NH_2$
Methionine	met	M	$-CH_2-CH_2-S-CH_3$
Phenylalanine	phe	F	$-CH_2$-phi
Proline	pro	P	$-[N]-(CH_2)_3-[CH]$
Serine	ser	S	$-CH_2-OH$
Threonine	thr	T	$-CH-(CH_3)-OH$
Tryptophan	try	W	$-CH_2$-indole
Tyrosine	tyr	Y	$-CH_2$-phi-OH
Valine	val	V	$-CH-(CH_3)_2$

Genetic Code

```
TTT phe F TCT ser S TAT tyr Y TGT cys C
TTC phe F TCC ser S TAC tyr Y TGC cys C
TTA leu L TCA ser S TAA OCH Z TGA OPA Z
TTG leu L TCG ser S TAG AMB Z TGG try W

CTT leu L CCT pro P CAT his H CGT arg R
CTC leu L CCC pro P CAC his H CGC arg R
CTA leu L CCA pro P CAA gln Q CGA arg R
CTG leu L CCG pro P CAG gln Q CGG arg R

ATT ile I ACT thr T AAT asn N AGT ser S
ATC ile I ACC thr T AAC asn N AGC ser S
ATA ile I ACA thr T AAA lys K AGA arg R
ATG met M ACG thr T AAG lys K AGG arg R

GTT val V GCT ala A GAT asp D GGT gly G
GTC val V GCC ala A GAC asp D GGC gly G
GTA val V GCA ala A GAA glu E GGA gly G
GTG val V GCG ala A GAG glu E GGG gly G
```

ACKNOWLEDGEMENTS

First and foremost many thanks to both my supervisors Ron Hay and Jim Naismith, whose patience, encouragement and direction have provided a constant inspiration. Thanks also to my parents and friends for their continued support and encouragement throughout. I would also like to acknowledge any members of the research teams in the Biomolecular Sciences Building (past and present) in St. Andrews, who have been kind enough to help me in whatever way during my studies. Finally, I would like to make a general acknowledgement to all those who have been involved in my education whether scientific or otherwise, from school to university. I feel very fortunate to have been taught and cultured by such dedicated people.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC).

ABSTRACT

Dynamic post-translational modifications are ubiquitous and diverse cellular tools for reversibly altering specific characteristics of the proteins to which they are conjugated. The Small Ubiquitin-like Modifiers (SUMO-1, SUMO-2 and SUMO-3) belong to a family of ubiquitin-like proteins (UBLs) which are covalently attached to, and removed from specific protein substrates via distinct pathways parallel to those of ubiquitin (Ub). Unlike ubiquitination which generally targets proteins for degradation, modification with SUMO-1, the most heavily researched of the SUMO proteins, appears to influence sub-cellular partitioning and protein stability, while protein modification by SUMO-2/-3 appears to be stress related and is of unknown consequence.

To help clarify distinctions between the dynamic SUMO-1-modifying system, and the systems of Ub and other UBLs, the NMR structure of SUMO-1 was investigated. Although sharing only 18% amino-acid sequence identity with Ub, the tertiary structure of SUMO-1 closely resembles that of Ub, containing the ββαββαβ fold characteristic of the Ub protein family. However, in contrast with Ub, SUMO-1 presents a generally negative electrostatic potential, with many regions known to contain basic residues in Ub, being substituted by acidic amino-acids in SUMO-1. One such region of difference is that corresponding to residues 93-95 in SUMO-1 (72-74 in Ub) adjacent to the C-terminal diglycine involved in the covalent attachment to target proteins. C-terminal mutants of SUMO-1 in which residues Glu93, Gln94 and Thr95 were substituted for the corresponding residues in Ub, Arg72, Leu73 and Arg74,

revealed the importance of these amino-acids for efficient metabolism by SUMO-1-specific conjugating and proteolytic enzymes.

Like SUMO-1 and Ub themselves, comparisons between their respective conjugating E2 enzymes reveals a combination of striking structural similarity with significant surface charge anomalies. In particular, a polar two residue Asp100-Lys101 insertion in the SUMO-1-specific E2 Ubc9, known not to be conserved in other E2 enzymes, impinges on the catalytic cysteine residue producing unique localised charges close to the active site of the protein. Functional analysis of site-directed mutants of this insertion suggested that these residues may be involved in recognition of the SUMO-1 substrate consensus motif.

SUMO-2 and SUMO-3 have 95% sequence identity between themselves, and 50% with SUMO-1. To determine the enzymes responsible for their conjugation, cDNAs for SUMO-2 and SUMO-3 were cloned from purified HeLa cell mRNA using RT-PCR into expression vectors, and the proteins functionally analysed. Although their respective profiles of conjugated proteins *in vivo* differ from those of SUMO-1, no discernible difference in *in vitro* conjugation activity could be seen, suggesting that elements other than the SUMO-specific E1 and E2 enzymes alone are responsible for the observed *in vivo* distinctions.

1. INTRODUCTION

1.1. Reversible Enzymatic Post-translational Modifications

In all living species the ribosomal translation of messenger RNA into the newly formed polypeptide or protein using the 19 codon-specified amino-acids (and one imino-acid), is by no means the final word in the structure, function or cellular localisation of that protein. Indeed, the properties of many, if not all cellular proteins are thought to be altered by modifications of one type or another between the events of their synthesis and degradation. A huge variety of protein modifications are known, ranging from the formation of intra-protein chemical bonds to the cleavage of the protein chain itself, but the majority involve the covalent attachment of molecular groups to specific residues of the protein. In fact, prior to synthesis (pre-translationally), during synthesis (co-translationally) and after synthesis (post-translationally) the characteristics of many cellular proteins can be subtly or drastically altered by the covalent modification of amino-acid residues, to generate "secondary amino-acids". The frequency and variety of these modifications are not insignificant, and although the total number of secondary amino-acids is unknown, they are thought to number approximately 200 (Krishna and Wold, 1993). Modifying groups range enormously in size, from small di-atomic species of a few atomic mass units, to large multi-subunit moieties with molecular masses in the 10's of thousands. Groups can comprise small organic or inorganic molecules, proteins, lipids, sugars, vitamins, and even combinations of individual molecular types in a single molecule. The most salient forms of amino-acid modifications are known to be; (a)

hydroxylation of proline and lysine, (b) phosphorylation of serine, threonine, tyrosine, arginine and histidine, (c) sulfation of tyrosine, (d) adenylation of tyrosine, (e) carboxylation of glutamate, (f) amidation at C-terminal residues, (g) acetylation of lysine, (h) methylation of lysine and arginine, (j) ubiquitination of lysine, (k) glycosylation of arginine, lysine, serine, threonine, and tyrosine and (l) lipidation of cysteine, serine and threonine residues (Han *et al.*, 1991).

The reversible enzymatic post-translational modifications are in general the most highly regulated and dynamic of all protein modifications. Their cellular significance cannot be understated. Most of them are singularly essential for normal cellular and physiological functioning, and together are thought to influence every known cellular process (Krishna and Wold, 1993). However, involvement in such a wide variety of important mechanisms also carries with it the responsibility that should such systems become defective, they often lead to disease states. In particular aberrations in modification pathways have implications upon numerous significant diseases, including Rheumatoid arthritis, Alzheimer's diseases, diabetes and cancer (Archer and Hodin, 1999; Buee *et al.*, 2000; Dennis *et al.*, 1999; Franz *et al.*, 2000; Kornfield, 1998; Patarca, 1996; Schwartz and Ciechanover, 1999), to name only a few.

Although each enzyme-catalysed post-translational modification can be considered as an individual system in its own right, the present understanding of cross-talk, mutual exclusivity and inter-dependency, is continually increasing in variety and complexity. Good examples of how many modification systems converge on single important cellular

molecules include the chromatin structure defining proteins histones, and the tumor suppressor proteins c-Jun and p53. All are known to be modified by a number of different species such as acetyl, phosphoryl, methyl, N-acetylgluosaminyl groups and Ub, sometimes on the same residues, and often with subtly varying or contrasting effects. Clearly the field of post-translational protein modifications is enormous and complex. This introduction will summarise some of the reversible enzymatic post-translational modifications which are either heavily researched today or predicted to be major players in the regulation of cellular processes. The summary will use examples of small inorganic (O-phosphorylation), small organic (N-acetylation), carbohydrate (O-linked N-acetylglucosaminylation) and lipid (S-palmitoylation), but will focus mainly on the protein modifying groups of Ub and the ubiquitin-like protein modifiers (UBLs). Table 1 gives a brief outline of the basic information known for each modification.

1.2. O-phosphorylation

Protein phosphorylation has been heavily researched since the first protein kinase (phosphorylase kinase) was isolated in 1959 (Krebs *et al.*, 1959). Phosphorylation is generally regarded as the most common and ubiquitous form of covalent protein modification. The most commonly phosphorylated residues are Ser, Thr and Tyr via hydroxyl groups, by a reversible process known as O-phosphorylation. Basic amino-acids such as Arg, Lys and His are also known to be phosphorylated on guanido and

MODIFICATION (Chemical Nature of Group)	STRUCTURE (M_r/Da)	MODIFIED RESIDUES (Enzymes of Conjugation & Deconjugation)	IDENTIFIED SUBSTRATE(S)	COMMON SUBSTRATE SIGNIFICANCE OF MODIFICATION	KNOWN REGULATED CELLULAR PROCESSES
O-phosphorylation (Small inorganic)	$-PO_3^{3-}$ (78)	Ser, Thr & Tyr (Kinases & phosphatases)	Enormous spectrum of proteins from all cellular locations	Modulation of structure, activation, protein:protein interactions. Signal for/blocks further modifications.	Involved in all major cellular processes.
N-acetylation (Small organic)	$-COCH_3$ (43)	Lys (Acetyltransferases & deacetylases)	Histones, nuclear import proteins and transcription factors.	Modulation of protein structure, stability and protein:protein interactions.	Gene transactivation, chromatin condensation. Cell-cycle control. Nuclear import.
O-linked N-acetylglucosaminy-lation (Sugar)	$-C_8NH_{14}O_6$ (220)	Ser & Thr (N-acetyglucosaminyltransferase & β-D-N-acetylglucosaminidase)	Nuclear pore, cytoskeletal and chromatin proteins, transcription factors, oncoproteins & Tyrosine phosphatases.	Modulation of activity, protein:protein interactions, stability, structure. Blocks phosphorylation.	Signal transduction, gene transactivation, nuclear import, cell-cycle control, cytoskeletal assembly.
S-palmitoylation (Lipid)	$-C_{16}H_{31}O$ (289)	Cys (Palmitoyltransferase & palmitoylthioesterase)	Membrane proteins (G-protein α-subunits, ion channels, neurotransmitter receptors & cell adhesion components).	Modulation of protein:protein interactions, activation. Membrane targeting.	Cell response to external stimuli, surface membrane trafficking, synaptic function.
Ubiquitination (Protein)	76 residue polypeptide (8559)	Lys (Ub-specific E1, E2s and E3s & Ub-isopeptidases)	Huge number of proteins, mostly short-lived/misfolded.	Targets (mostly) for 26S proteasome-mediated degradation.	Influences a huge variety of major cellular processes
NEDD8 conjugation (Protein)	76 residue polypeptide (8554)	Lys (NEDD8-specific E1, E2 and E3 & NEDD8 isopeptidases)	Ubiquitin E3 scaffold proteins, cullins.	Modulates protein:protein interactions and stabilisation.	Cell-cycle progression. Gene transactivation.
SUMO-1 conjugation (Protein)	97 residue polypeptide (11125)	Lys (SUMO-1-specific E1, E2 and E3? & SUMO-1 specific isopeptidases)	Nuclear dot associated proteins, nuclear pore proteins, surface receptors, p53, IκBα etc	Protein stabilisation, sub-cellular localisation and activation state.	Gene transactivation, nuclear import, sub-nuclear structural integrity, signal transduction.

Table 1. General summary of examples of small organic-, small inorganic-, sugar-, lipid- and protein-based reversible enzyme-catalysed post-translational modifications with the capacity to regulate important cellular processes. "?" indicates no *in vitro* but possible *in vivo* existence.

imidazole groups, termed N-phosphorylation, but other than the reactions catalysed by the bacterial 'relay histidine kinases (Dutta *et al.*, 1999), this process is thought not to be dynamically regulated (Han *et al.*, 1991). Reversible phosphorylation has implications on numerous essential physiological and cellular processes. Some of the most heavily studied processes include meiosis (Dekel, 1996), apoptosis (Gjertsen and Doskeland, 1995), cell signalling (Fischer, 1999), neuronal transmission and membrane ion shuttling (Levitan, 1999; Swope *et al.*, 1999), signal transduction (Barik, 1996), endo- and exocytosis (Liu, 1997) and cell cycle control (Meek *et al.*, 1997). Consequently great academic and industrial interest has been generated in the mechanisms and pathways which regulate the phosphorylation state of important cellular proteins. The enzymes responsible for protein O-phosphorylation, the protein O-phosphorylation protein kinases (referred to herein as "kinases"), have classically been divided into two groups, Ser/Thr kinases and Tyr kinases, reflecting the residue specificity of the enzymes. Ser/Thr kinases are often distinguished by their mode of regulation, e.g. cyclic nucleotide-dependant, calmodulin-dependent, diacylglycerol and calcium-regulated protein kinase C (Edelman *et al.*, 1987; Hunter, 1987). Tyrosine kinases, on the other hand, are commonly known to be cell membrane anchored receptors, such as the epidermal growth factor (EGF), platelet derived growth factor (PDGF) and insulin receptors, which become activated in response to ligand binding (Hunter and Cooper, 1985; Lau *et al.*, 1989; Moudgil, 1989). Especially important to the role of phosphorylation in signal transduction, is the fact that phospho-tyrosine residues are

recognised by SH2 (src homology region 2) domains of target proteins, which allows the transmission and amplification from signal to effect (Stryer, 1995).

The huge role of phosphorylation as a regulatory modification is reflected in the fact that most kinases can themselves be regulated positively or negatively by phosphorylation. Sequence comparisons of all Ser/Thr and Tyr kinases reveal similarity in the catalytic domain, thus allowing prediction of whether any protein sequence may be a kinase (Hunter, 1987). Substrate recognition sites have been examined, and although no single motif exists for each class of phosphorylation, groups of kinases are known to recognise specific motifs (Kemp and Pearson, 1990).

The hydrolysis of attached phosphoryl groups is catalysed by phosphatases. The study of protein phosphatases has been somewhat less intense than that of kinases, but their regulatory significance is thought to be as important. They appear to be far fewer in number and the majority have a much broader specificity range than kinases. Protein phosphatase 2A (PP2A), for example is involved in a wide range of systems including glycolysis, lipid metabolism, and catecholamine metabolism (Cohen *et al.*, 1990). Of all the reversible enzymatic post-translational modifications, phosphorylation/dephosphorylation has been studied most intensely, and is generally used as the yardstick against which all other modification systems are compared.

1.3. N-acetylation

It has been suggested that about 80% of the soluble proteins in mammalian cells have acetylated N-terminal amino-acids (Wellner *et al.*, 1990). This type of acetylation is, however, constitutive and irreversible, and although it is thought to play a limited role in protein protection from N-terminal proteolysis, its specific functions are poorly understood. In contrast, intra-chain lysine acetylation (N-acetylation, but referred to herein simply as acetylation) is a dynamic cellular process with specific and subtle cellular significances. The transfer of acetyl groups from acetyl-CoA onto lysine residues of target proteins is catalysed by acetyl-transferases (see Figure 1 for schematic). The major targets of acetylation are the chromatin structure defining proteins, histones. In eukaryotes DNA typically exists *in vivo* as a repeating array of nucleosomes in which 146bp of DNA are wrapped around a histone octamer consisting of pairs of histone proteins H2A, H2B, H3 and H4 (Wolffe, 1992). Such organisation is repressive to gene transcription by making the DNA unavailable for the binding of transcriptional machinery and other regulatory factors. Over 30 years ago the acetylation of histones was proposed to be associated with the transcriptional activity of local regions of DNA in eukaryotic cells (Allfrey *et al.*, 1964; Pogo *et al.*, 1966). Histones are acetylated by histone acetyl-transferases (HATs) at specific N-terminal lysine residues, which is thought to disrupt interactions with DNA, and open chromatin (Bayle and Crabtree, 1997). HATs and the acetyl deconjugating enzymes, histone deacetylases (HDACs) can thus be considered as transcription coactivators and corepressors respectively, which transduce extracellular signals

(Magnaghi-Jaulin *et al.*, 1999; Sterner and Berger, 2000). Until recently the dogma had been that chromatin structure and modification are independent of transcriptional initiation, although that has since been disproved. Proteins which contain intrinsic histone acetylase and deacetylase activity have been identified as components of the RNA polymerase II transcription machinery, proteins that associate with transcriptional regulatory factors and proteins that positively or negatively affect transcription *in vivo* (Struhl, 1998). Essentially the specific recruitment of HATs or HDACs to promoters is the generally accepted model for transcriptional regulation through histone acetylation (Bayle and Crabtree, 1997).

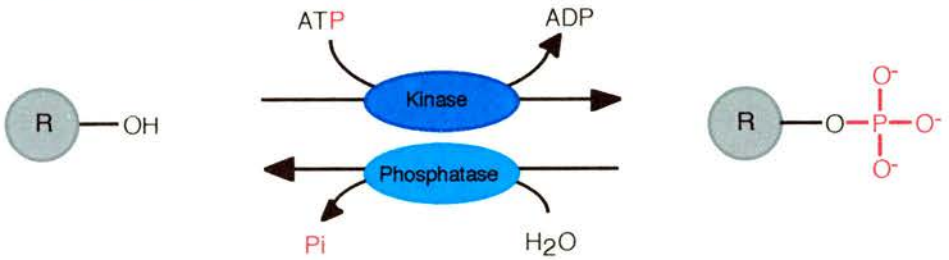
Protein N-acetylation is now known not to be limited solely to histones. Recently other non-histone proteins have been shown to be reversibly acetylated, including nuclear import factors (importin α), the microtubule component α-tubulin, and even other acetylases (P/CAF and P300), with effects on DNA binding, protein stability and protein-protein interactions (Kouzarides, 2000). However, the acetylation of these proteins appears not to be catalysed by dedicated acetylases, but those previously shown to acetylate histones. Specifically, the tumor suppressor protein p53 is known to be acetylated. p53 performs its tumour suppressor role as a transcription factor which binds to specific DNA sequences upstream to many genes whose protein products regulate cell cycle progression and apoptosis (Ko and Prives, 1996). It is known to be acetylated on two lysine residues within clusters of basic residues in the C-terminus by the HAT p300, which activates its DNA binding ability (Gu and Roeder, 1997).

Somewhat confusingly, due to the substrate diversity of some HATs for non-histone proteins, the literature continues to refer to them as histone acetyl-transferases, although pressure has risen to refer to such enzymes as factor acetyl-transferases (FATS) (Sterner and Berger, 2000). The huge variety of enzymes in the nucleus and cytoplasm with known HAT or HAT/FAT activity, has meant that no single model of a typical enzyme exists. Although they vary in size, in general they are monomers found in multi-protein complexes, and most have the ~150 residue "HAT domain" known to contain conserved residues essential for HAT catalytic activity (Sterner and Berger, 2000).

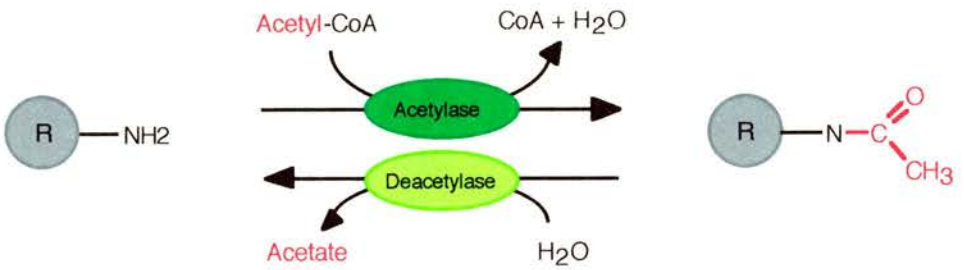
Although the present understanding of protein N-acetylation is relatively limited in comparison, it does stand up to analogy with phosphorylation, when the substrate diversity and varied functional significances are considered. Where comparisons with phosphorylation break down, is when one considers that unlike phosphorylation there is no apparent acetylation cascade i.e. an acetylase modifying the enzymatic activity of a second acetylase in order to transmit a biological signal (Kouzarides, 2000). However, the elements for the implementation of an acetylation cascade have been identified by the discovery of the bromodomain which recognises acetyl-lysines. The bromodomain is thus thought to be analogous to the SH2 domain of the tyrosine phosphorylation cascade (Kouzarides, 2000).

Figure 1. Schematic representation of conjugation and deconjugation reactions for O-phosphorylation, N-acetylation, O-linked N-acetylglucosaminylation and S-palmitoylation. R represents an acceptor protein. O-phosphorylation occurs on hydroxyl groups, catalysed by Ser/Thr/Tyr protein kinases. N-acetylation on ε-nitrogens of Lys residues is catalysed by acetyl-transferases. O-linked N-acetylglucosaminylation on hydroxyl groups of Ser, Thr and Tyr residues is catalysed by N-acetyglucosaminyl-transferase (O-GlcNAc-T). S-palmitoylation occurs on the sulfur groups of Cys residues, catalysed by palmitoyl-transferases (Palmitoyl-Tr). Deconjugation reactions involve phosphatases, deacetylases, β-D-N-acetylglucosaminidase (O-GlcNAcase), and palmitoyl-protein-thioesterases (Palityl-Th or PPTs) respectively. Figure compiled from (Aletta *et al.*, 1998; Casey, 1995; Hart, 1997; Sterner and Berger, 2000; Stryer, 1995).

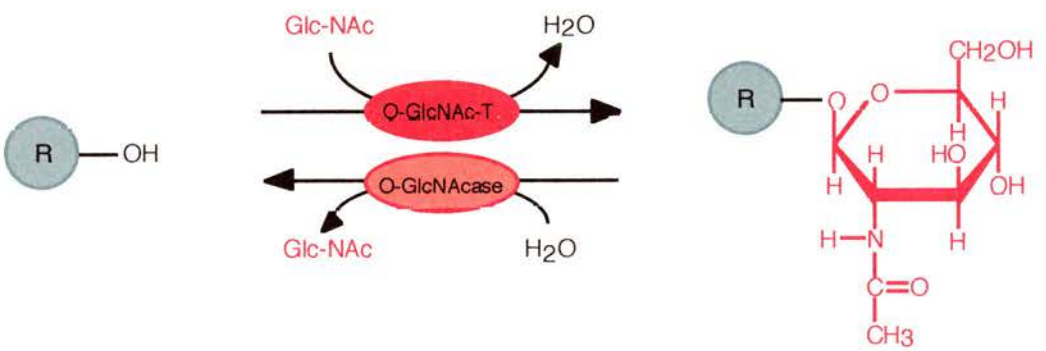
O-phosphorylation



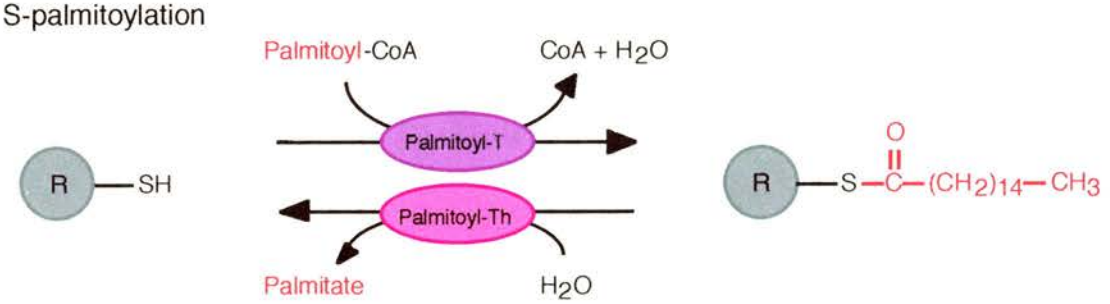
N-acetylation



O-linked-N-acetylglucosamiylation



S-palmitoylation



1.4. O-linked N-acetylglucosaminylation

It is thought that as many as two-thirds of all proteins in biological systems are glycosylated (Apweiler *et al.*, 1999). In contrast to the small modifiers discussed above, it is often the case for glycoproteins that the molecular volume occupied by the oligosaccharide is larger than that of the protein itself. The oligosaccharide components of glycoproteins are linked to either Asn side-chains by N-glycosidic bonds, or Ser/Thr side-chains by O-glycosidic bonds. Approximately 90% of glycoproteins contain N-linked glycosylations only or N- and O-linked types, while the remaining 10% are O-linked only (Apweiler *et al.*, 1999). N- and O-linked oligosaccharides are processed in the endoplasmic reticulum (ER) and the *cis, medial* and *trans* compartments of the Golgi. As such N- and O-linked glycosylation in the ER and Golgi is both co- and post-translational (Imperiali and O'Connor, 1999). Large branched polysaccharides containing mannose, galactose, glucose, fucose and arabinose sugars, are generated in the Golgi complex which subsequently targets these proteins to lysosomes, secretory vesicles, and the plasma membrane. These linked oligosaccharides are known to influence protein aggregation, stability, and protease and heat resistance, thus indicating their significance in protein activity and function.

Until a few decades ago the dogma had been that glycoproteins only existed in the lumenal compartments of the cell. However, since then many forms of nuclear and cytoplasmic glycosylation have been identified

(Hart *et al.*, 1989). Of particular significance is the O-linked N-acetylglucosaminylation (O-GlcNAcylation), a relatively recently discovered form of intracellular glycosylation. The modification is characterised by a single, unmodified N-acetylglucosaminyl (GlcNAc) moiety linked through a glycosidic bond to the hydroxyl groups of Ser/Thr residues (shown schematically in Figure 1), often occurring on multiple sites of the same molecule (Snow and Hart, 1998). Unlike the classical forms of complex and Asn-linked glycans existing in the secretory pathway, O-GlcNAc-containing proteins appear to be glycosylated and deglycosylated in a dynamic fashion (Hart, 1997). To date only single examples of the conjugating and deconjugating enzymes have been identified. The conjugation reaction is catalysed by N-acetylglucosaminyl-transferase (O-GlcNAc-transferase), a heterodimeric enzyme consisting of a 110 kDa catalytic α-subunit and a 78 kDa β-subunit of as yet unknown function. The deconjugation reaction is catalysed by β-D-N-acetylglucosaminidase (O-GlcNAcase), again a heterodimer of α/β subunits with the 54 kDa α-subunit containing the active site and the 51 kDa β-subunit of unidentified role (Hart, 1997). O-GlcNAc-modified proteins have been found in all eukaryotic cells examined (Kearse and Hart, 1991). Known O-GlcNAc-modified proteins are functionally diverse, yet they have two common features. Firstly, they are known also to be targets for phosphorylation, and secondly, they form reversible multi-meric complexes with other polypeptides via association, that are regulated by phosphorylation (Hart, 1997). The highest concentration of O-

GlcNAcylated proteins occurs in the nuclear pore complex (NPC) (Holt *et al.*, 1987), yet the greatest abundance are found in the nucleus, and to a lesser degree, in the cytoplasm (Hart *et al.*, 1988). Proteins known to be modified in eukaryotic cells include nuclear pore proteins (p62, p58 and p54), cytoskeletal and membrane proteins (intermediate filament, microtubule associated, and cytoskeletal bridging proteins) and proteins from infectious organisms (viruses and parasites) (Hart, 1997). But it is the role of O-GlcNAcylation in oncogenesis which has attracted the greatest research activity recently. Again, familiar proteins appear in the lists of oncogenic targets for modification, such as p53, c-Myc and the SV40 large T-antigen. Like acetylation, O-GlcNAcylation of p53 occurs at its basic C-terminus (Shaw *et al.*, 1996). Although the effect of the modification has not been characterised it is likely that as with other C-terminal p53 modifications, it plays a role in modulating the affinity of the protein for DNA. On the other hand, c-Myc has been thoroughly examined for its conjugation with GlcNAc. c-Myc is a phosphoprotein that heterodimerises with the Max protein and regulates gene transcription. The O-GlcNAcylation occurs on Thr58 within the transactivation domain, which binds the retinoblastoma gene product RB, and is a mutational hot-spot in Burkitt's and other lymphomas (Chou *et al.*, 1995). Thr58 is also a site of phosphorylation in c-Myc, and mutation of Thr58 results in an increased tumor-inducing potential in animals. This suggests that the reciprocal glycosylation and phosphorylation of Thr58 may play an important role in modulating the functions of c-Myc.

Like acetylation O-GlcNAcylation has been compared with phosphorylation. In fact it has been suggested that it is as abundant and dynamic as phosphorylation itself on many eukaryotic nuclear and cytoplasmic proteins (Hart, 1997). Unfortunately, given this proposition and the fact that Ser/Thr-O-GlcNAcylation was discovered over 15 years ago (Torres and Hart, 1984) relatively little is known about the functions and regulation of the process. As such, whether or not they play a key role in cellular processes has yet to be clearly demonstrated.

1.5. S-palmitoylation

In the past 15 years pathways have been highlighted that involve co- and post-translational modification of proteins by specific lipids, generally termed lipidation (Casey, 1995). Lipid modified proteins are classified on the basis of the identity of the attached lipid. The common outcome of protein lipidation is the direction of the modified species to cellular membranes. Four major classes of lipidation have been identified; myristoylation, palmitoylation, prenylation and modification with the complex glycosylphosphatidylinositol (GPI) moiety. Myristoylation is a co-translational procedure, generally regarded as a constitutive process, resulting in a stably modified protein, targeted to the cytoplasmic face of the lipid bilayer (Casey, 1995). GPI, as the name suggests is an entire phospholipid with associated sugars and ethanolamine. GPI-modified proteins obtain the moiety post-translationally in the ER, and essentially all GPI-modified proteins are destined for the cell surface. Palmitoylation

and prenylation differ from the other two classes of lipidation as both post-translational processes are attached via a thioester bond to cysteine residues in the protein. Like myristoylation, prenylation is a constitutive process and results in the generation of a stably modified protein, although in contrast, the post-translational modification of cysteine residues with palmitate occurs through a labile thioester bond by a process known to be reversible (see schematic of conjugation reaction in Figure 1). This is apparently unique for a lipid modification (Casey, 1995). To date, identified S-palmitoylated proteins are mostly membrane bound and almost entirely found in neural cells (Bizzozero, 1997). Examples include ion channels, neurotransmitter receptors, signal transduction components and cell adhesion molecules. Based on such studies, the primary function of protein palmitoylation like all lipidations, is thought to be membrane targeting, although it can also facilitate increased specific protein-protein interactions and structural activation (Bizzozero, 1997; Wedegaertner *et al.*, 1995). Two enzyme activities which catalyse the attachment of palmitoyl groups, palmitoyl-transferases, have been identified, although their activities have not been fully characterised (Berthiaume and Resh, 1995; Dunphy *et al.*, 1996; Kasinathan *et al.*, 1990). Palmitoyl-transferases have not been cloned to date, but are known to be approximately 65 kDa in mass and localise predominantly to the plasma membrane, consistent with the cellular locality of their substrates.

S-palmitoylation is unique among the reversible post-translational modifications, as the majority of research in the field has been directed at the proteins responsible for the removal of the conjugated moiety, as

opposed to the attachment. Palmitate deconjugation is mediated by palmitoyl-protein thioestérases (PPTs) of which two activities, PPT1 and PPT2, have been cloned and analysed to date (Camp *et al.*, 1994; Soyombo and Hofmann, 1997). The reason for this research bias is due to the identification of defects in the PPT1 gene in children suffering from infantile neuronal ceroid lipofuscinosis INCL (Vesa *et al.*, 1995), a fatal disease characterised by the accumulation of storage bodies in neural tissues. PPT1 is a 37 kDa protein, predominantly localised to lysosomes (Verkruyse and Hofmann, 1996) and has been shown to be developmentally regulated in brain and other neural tissues, but not in non-neural cell-types (Suopanki *et al.*, 1999). Conversely, PPT2 has only been partially characterised (Soyombo and Hofmann, 1997). PPT2 is 18% identical to PPT1 and is also found in lysosomes, yet is known to have a distinct substrate specificity and is thought not be involved in the pathogenesis of (INCL) (Soyombo and Hofmann, 1997). With the current bias of research, it is difficult to assess the overall cellular implications of reversible lipid modification, but it seems unlikely to be as ubiquitous a modification as the other major regulators discussed here.

1.6. Ubiquitination

The steady-state levels of cellular proteins is maintained by a delicate balance between synthesis and degradation. Specific proteins are degraded with half-lives that vary from more than 120 days (haemoglobin), to 10 min (c-Myc). At present two cellular systems for the

degradation of proteins have been identified; the vacuolar pathway (including lysosomes, and endosomes), and the cytoplasmic Ub-mediated pathway. Ub is a highly conserved 76-amino-acid protein that was first isolated over 25 years ago (Goldstein *et al.*, 1975) and exists in either a covalently linked, or free state. The Ub-mediated protein degradation pathway comprises two discrete and successive steps. The first involves the covalent attachment of multiple copies of Ub to the protein target, which directs the protein to the 26S proteasome (multicatalytic protease complex), where it is degraded in the second phase (Ciechanover and Schwartz, 1998). The Ub-protease system is the principal mechanism for the turnover of short-lived proteins, and is also involved in the selective degradation of abnormal and mutated proteins, as well as the processing of major histocompatibility complex (MHC) class I-restricted antigens (Ciechanover and Schwartz, 1998). The number of known substrates is large and extremely diverse. They include regulators of cell-cycle division such as G1 and mitotic cyclins and cyclin-dependant kinase inhibitors, growth regulators like c-Fos and c-Jun, tumor suppressors such as p53, surface receptors like the growth-hormone receptor, ion channels such as the cystic-fibrosis transmembrane conductance regulator (CFTR), and transcription regulators like IκBα (Ciechanover and Schwartz, 1998; Hay *et al.*, 1999; Schwartz and Ciechanover, 1999). The physiological consequence of such a diverse substrate specificity is that Ub-mediated protein degradation influences a huge spectrum of basic cellular processes, including regulation of cell cycle and division, cell response to stress/extracellular modulators, cell morphogenesis, neuronal networks

activity, modulation of cell surface receptors, ion channels, DNA repair, biogenesis of organelles and the regulation of immune and inflammatory responses (Ciechanover and Schwartz, 1998; Schwartz and Ciechanover, 1999).

1.6.1. The ubiquitin conjugation pathway

The conjugation of Ub to protein substrates generally requires the activity of three enzyme species, and follows a three step conjugation mechanism (as shown in Figure 2). In the first step Ub is activated via an adenylate intermediate, by the ATP dependent formation of a thioester bond between the carboxyl group of the C-terminal Gly (76) of Ub and a cysteine residue of a Ub-activating enzyme (E1). This is then transferred in the second step, to a cysteine residue within a Ub-conjugating enzyme (E2) via a transesterification reaction. Finally Ub is transferred to the protein acceptor by the generation of an isopeptide bond between the carboxyl group of Gly76 and ε-amino group of a specific lysine residue. In many cases the final step requires the participation of a Ub protein ligase (E3) which acts as either the ultimate Ub donor or as the substrate recognition factor (Hershko and Ciechanover, 1998).

To date only one Ub-activating enzyme E1 has been identified, UBA1, which exists as two isoforms of 110 kDa and 117 kDa. They are found both in the cytoplasm and nucleus, and deletion of the one encoding gene is known to be lethal (Schwartz and Ciechanover, 1999). A large family of Ub-conjugating E2 enzymes have been identified in a

number of eukaryotic species, totalling over 100 sequences to date (over 20 from mammals). They are generally small in comparison with E1 enzymes (14-35 kDa), but similarly are found both in the cytoplasm and nucleus. The activity of a number of E2 enzymes have been characterised, and individual E2s can be assigned roles in the regulation of specific cellular processes (Hochstrasser, 1996). The absolute substrate specificity of E2s is largely due to their association with particular E3 enzymes, which bind specific protein substrates. Furthermore it is thought that different E2s do not interact with only one E3 enzyme (or vice-versa), but that a certain degree of flexibility exists. For example, the ubiquitination of proteins catalysed by the E3 enzyme E6-AP, probably involves the E2s Ubch5 or Ubch5B, Ubch5C as well as Ubch7 or Ubch8 (Schwartz and Ciechanover, 1999). As mentioned, the nature of the actual transfer reaction catalysed by different E3s is varied. In some cases, an E3 accepts activated Ub from an E2 and binds it as a thioester intermediate prior to transfer to substrate, while in other cases the E3 acts as an associating protein which brings E2 and substrate together and transfer passes directly from E2 to protein (Tanaka *et al.*, 1998). To date, four distinct types of Ub-protein ligases have been identified; E3α or Ubr1, the HECT (homologous to E6-AP C-terminus) domain family, the "Destruction Box" or APC anaphase promoting complex/cyclosome E3s and the SCF (Skp1, cullin, F-box) family (Schwartz and Ciechanover, 1999; Hershko and Ciechanover 1998; Tyers and Willems 1999). The 200 kDa E3α (the "N-end rule" E3) and its yeast homologue Ubr1p binds proteins which have either a bulky hydrophobic,

or a basic N-terminal amino-acid. E3α is also thought to bind unfolded proteins via a putative "body" site, and is known to transfer Ub from a specific E2, Ubc2 to its protein targets (Schwartz and Ciechanover, 1999). A second member of the group, E3β is related to E3α, but appears to be specific for uncharged N-terminal amino-acids (Heller and Hershko, 1990). The HECT (homology to E6-AP C-terminus) domain family of E3s are related to the 100 kDa E3 required by the papilloma virus E6 oncoprotein for the Ub-mediated degradation of p53, known as E6-associated protein (E6-AP). E6-AP performs an intermediary role in Ub transfer to p53 as it is the E6 protein itself which brings p53 and the E3 together. The so-called "destruction box" ligase recognises a nine residue degenerate "destruction box" motif in its target proteins. This E3 is contained within high molecular weight multi-protein complexes known as anaphase-promoting complexes (APCs) or the cyclosome. It is specific for cell-cycle regulatory proteins which are degraded at the end of mitosis including mitotic cyclins, some anaphase inhibitors, and spindle associated proteins (Sudakin *et al.*, 1995). These high molecular weight complexes (~1500 kDa) are known to be activated by phosphorylation, but the specific ligase active subunits are yet to be identified. The SCF family of ligases are good examples of the combinatorial control of E3 ligase specificity (reviewed by Tyers and Willems 1999). The F-box proteins act as adaptor subunits which recognise substrates through specific protein-protein interactions. F-box proteins link up to a core catalytic complex consisting of Skp1, Cdc53 (or CUL1 on metazoans), and the Ub E2 Cdc34. There are now known to be hundreds of F-box proteins in sequence databases, reflecting the host of

proteins targeted for degradation by SCF pathways. Known substrates include cyclins, CDK inhibitors, and transcriptional regulators such as IκB and β-catenin (Tyers and Willems 1999).

Due to the fact that Ub can in fact modify itself via Lys 48, successive reactions of ubiquitination result in the formation of branched polymeric (polyUb) chains. It has been proposed that the formation of polyUb chains requires a specific polyUb chain assembly factors (E4s), which do not to take part directly in the conjugation cascade or interact directly with the substrate, but may in fact be members of a family of regulatory factors which act in concert with E3s to increase the specificity of ubiquitination. (Koegl *et al.*, 1999). It is the resultant polyUb signal which is recognised by the 26S proteasome as a marker of degradation.

1.6.2. The 26S proteasome

The protease complex involved in Ub-mediated degradation is the 26S proteasome. It is composed of a core 20S catalytic unit, flanked on both sides by 19S regulatory proteasome complexes (19S-20S-19S), and is distinct from other cellular proteases by its dependence on ATP and Ub, and by its complex structure. The crystal structure of the yeast 20S proteasome has given insight into eukaryotic proteasome function (Groll *et al.*, 1997). The 20S unit is barrel-like in shape comprising four stacked rings each containing seven distinct subunits α_{1-7}, β_{1-7}, β_{1-7}, α_{1-7}. Each sub-unit has molecular mass of between 25-30 kDa, and the active sites are known to be on β1, β2 and β5 subunits. The α-subunits are thought to play an essential

role in stabilising the two-ring structure of the β chains, and may be involved in the binding with the 19S cap complexes. The 19S complexes are composed of at least 18 distinct proteins with a molecular mass range of between 25 and 110 kDa. This complex serves as a port of entry into the catalytic core and provides the different regulatory functions that are necessary to ensure the selective degradation of Ub-tagged substrates. These include a binding site for Ub chains, Ub recycling activity, and several ATPases in addition to its ability to regulate the different peptidase activities of the 20S complex (Ciechanover and Schwartz, 1998).

Substrates carrying many molecules of Ub are more efficiently recognised by proteasomes than those carrying few. A subunit of the 19S complex has been identified which in both mammals (S5a) and plants (MBP1) binds Ub monomers, but has a much higher affinity for polyUb chains. Interestingly, the proteolytic activity of the proteasome is now thought not to be restricted only to ubiquitinated substrates. Proteasome-mediated Ub-*in*dependent protein degradation has been demonstrated for ornithine decarboxylase (ODC), and some ER proteins, suggesting that the substrate specificity of the proteasome may be broader than first thought (Sommer and Wolf, 1997).

1.6.3. Ubiquitination not as a signal for proteasome-mediated degradation

It is now clear that not all ubiquitinated proteins are targeted for degradation by the proteasome. This is true mostly for cell surface mature

membrane associated proteins such as the growth hormone receptor (Pickart, 2000), where the conjugation of a single Ub molecule (monoUb) is still a signal for degradation, but via a different mechanism from that described above. MonoUb has been shown to signal endocytosis, leading to proteolysis in the vacuole or lysosome (Terrell *et al.*, 1998). However, in contrast to the general role of ubiquitination as a degradation signal, a number of substrates for mono- and polyUb-conjugation have been shown to remain stable. One well studied example of non-degraded ubiquitinated substrates are the previously described histone proteins. The core histones H2A, H2B and H3 are known to be reversibly ubiquitinated in the C-terminal regions (Lys119 in H2A and Lys120 in H2B). Ubiquitinated histones are associated with transcriptionally active DNA, and are not targeted for degradation. (Spencer and Davie, 1999). The C-terminal modification of core histone proteins is unique to Ub, as the other known modifications phosphorylation, acetylation and methylation, are all targeted to the exposed N-term regions. It is thought that ubiquitination of core histones results in alteration of nucleosome and/or higher order chromatin structure (Davie and Murphy, 1990).

Perhaps the most intriguing discovery in the heterogenaity of polyUb protein modification is that Ub chains can be assembled through lysine residues of Ub other than Lys48, some resulting in targeting to functionally distinct biological processes. In the most striking example yeast strains expressing K63R Ub (blocked for chain assembly via Lys63) show defects in post replicative DNA repair (Spence *et al.*, 1995), the stress response (Arnason and Ellison, 1994), mitochondrial DNA inheritance

(Fisk and Yaffe, 1999), endocytosis of some plasma membrane proteins (Galan and Haguenauer-Tsapis, 1997), and ribosomal function (Spence *et al.*, 2000). In addition these reports also suggest that Lys63-linked chains have signalling function that does not involve proteolysis. Furthermore, Lys29-linked chains have been shown to have functionally distinct roles (Koegl *et al.*, 1999), but the exact implications have yet to be elucidated. It is now generally accepted that mono and poly-Ub (and heterogeneous polyUb signals) can be considered functionally distinct. An interesting hypothesis considers that other lysine-targeting post-translational modifications may be involved in the functional modification of Ub itself. However, although the *in vitro* modification of Lys48 with other Ub-like species has been demonstrated (Whitby *et al.*, 1998), whether or not such modifications of Ub exist *in vivo*, is debatable.

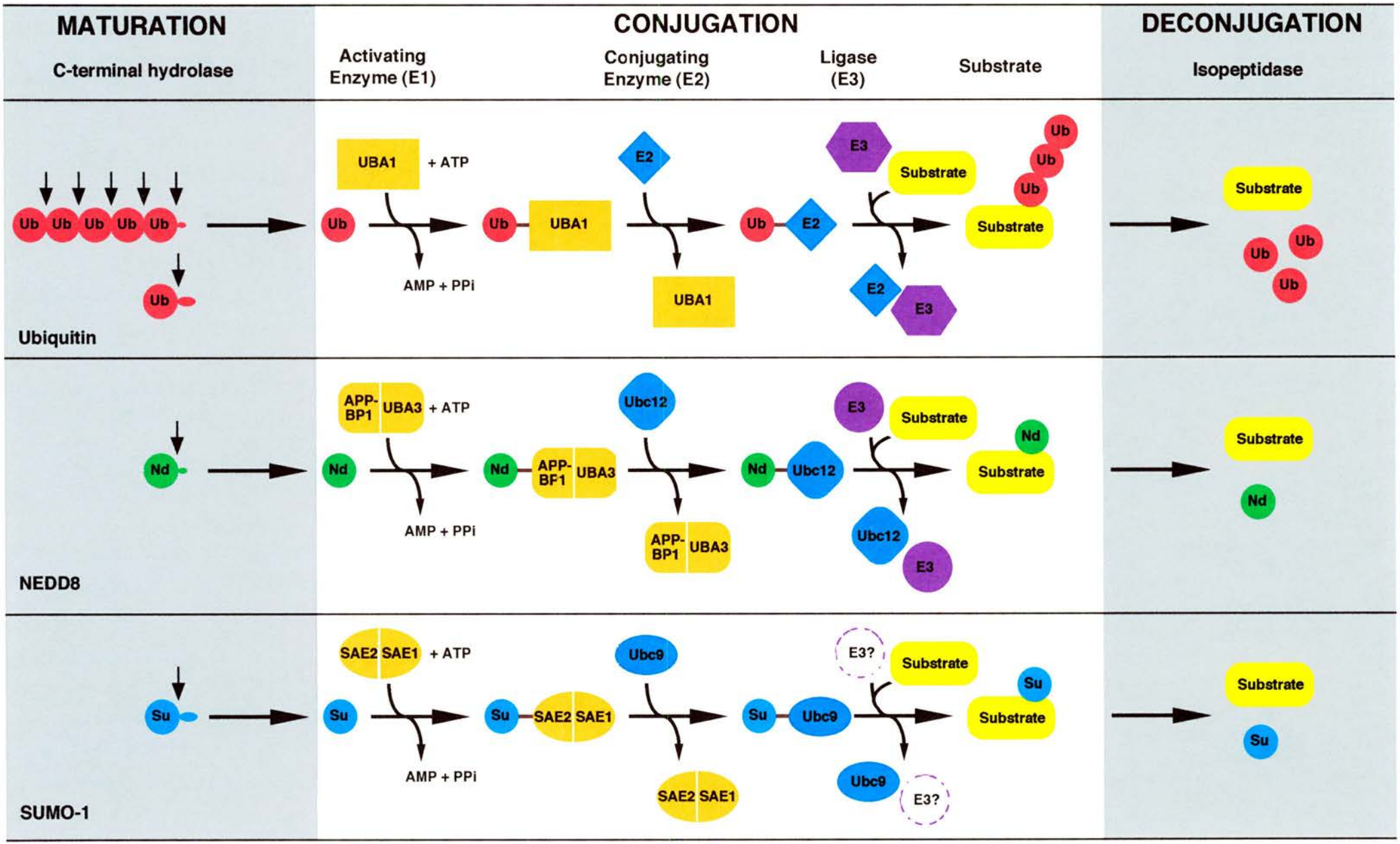
1.6.4. Deubiquitinating enzymes

Due to the intensity of the research surrounding the attachment of Ub to specific proteins targets, it is generally thought that the regulation of conjugation processes are the major controlling factors over the ubiquitination state of conjugated proteins. However, deubiquitination (deUb), the removal of Ub from proteins, is increasingly being recognised as an important regulatory strategy (D'Andrea and Pellman, 1998; Wilkinson, 1997b). DeUb is catalysed by processing proteases collectively known as deubiquitinating enzymes (DUBs). DUBs are cysteine proteases that cleave Ub-esters, Ub-thioesters or Ub-amides. DUBs act at multiple

levels in the Ub-proteasome pathway; processing of the primary Ub gene products (as shown in Figure 2, left hand side), salvage of Ub that has been trapped by reaction with small cellular nucleophiles, editing the ubiquitination state of proteins and disassembly of the polyUb signal (also shown in Figure 2, right hand side). More than 60 DUB sequences have been identified to date, making them the largest family of enzymes in the Ub system. They are known to be involved in numerous biologically important processes including control of growth, differentiation, oncogenesis, and genome integrity (Wilkinson, 1997a). Eukaryotes contain DUBs encoded by at least two gene families; the UCH family (Ub C-terminal hydrolase, also known as type I UCH), and the UBP family (Ub-specific processing proteases, also known as type 2 UCH) (Wilkinson, 1997a).

The UCH family of DUBs are known to process small peptides and protein domains from the C-terminus of Ub (Wilkinson, 1997b). Therefore they have a number of important roles in Ub metabolism, such as the removal of small peptide fragments from the polyUb released from the proteasome, salvage of trapped Ub, and the co-translational processing of the Ub gene products (Wilkinson, 1997b). All these functions are required to replenish the pool of free Ub molecules for (further) rounds of conjugation. However, at present the specific enzymes involved in each separate processing event, are largely unknown (D'Andrea and Pellman, 1998). A number of different UCH enzymes have been identified from different organisms, and sequence comparisons have revealed the existence of a common 210 amino-acid catalytic domain, along with four

Figure 2. Schematic diagram representing the maturation, conjugation and deconjugation reactions for Ub and the Ub-like modifiers NEDD8 and SUMO-1. The primary translation products for Ub (top), NEDD8 (Nd) (middle) and SUMO-1 (Su) (bottom) require removal of an inhibitory C-terminal extension before they can be attached to substrates (cleavage sites indicated by arrows). This maturation is catalysed by specific proteases known as C-terminal hydrolases. The cleaved modifiers are attached to protein substrates by a three step mechanism. In the first step specific activating enzymes (E1s) form ATP-dependent thioester bonds (shown in red) between internal cysteine residues and the conserved C-terminal glycine of the modifiers. The Ub E1 UBA1 is monomeric, while those of NEDD8 and SUMO-1 are heterodimers, APP-BP1/UBA3 and SAE1/SAE2 respectively. In the second step specific conjugating enzymes (E2s) accept the activated modifier via a transesterification reaction to an internal cysteine residue. Many E2s are known to exist for Ub (Ubc1-8, Ubc10, Ubc11 and Ubc13), while only single E2s have been isolated for NEDD8 and SUMO-1 (Ubc12 and Ubc9 respectively). The final step involves the formation of an isopeptide bond between the ε-amino group of the substrate target lysine and the terminal glycine of the modifier. This reaction is known to require an additional ligase protein (E3s) for Ub and to a lesser extent NEDD8 conjugation, but for SUMO-1 evidence of similar enzymes has not been conclusively demonstrated. The deconjugation reactions are catalysed by specific isopeptidases which cleave the isopeptide bond between the substrate and the modifier. Figure adapted from (Jentsch and Pyrowolakis, 2000; Yeh *et al.*, 2000).


MATURATION
C-terminal hydrolase
CONJUGATION
Activating Enzyme (E1)
Conjugating Enzyme (E2)
Ligase (E3)
Substrate
DECONJUGATION
Isopeptidase
Ub
UBA1
+ ATP
AMP + PPi
E2
E3
Substrate
Ubiquitin
Nd
APP-BP1
UBA3
Ubc12
NEDD8
Su
SAE2
SAE1
Ubc9
E3?
SUMO-1

other highly conserved blocks of sequence (Wilkinson, 1997a). Like all DUBs, the UCH enzymes contain a conserved catalytic triad of cysteine, aspartate and histidine residues (D'Andrea and Pellman, 1998). In general, the UCH enzymes are smaller (~30 kDa) and more closely related, than the UBPs (Larsen *et al.*, 1996).

The disassembly of polyUb chains or the editing of the ubiquitination state of proteins is catalysed by the second class of DUBs, the UBPs. Further to this activity, UBPs are also known to be able to process the primary Ub gene products in a manner analogous to UCHs. Sequence comparisons reveal that UBPs are more distantly related than the UCH enzymes, but they are characterised by the presence of six conserved regions spanning 300 to 500 amino-acids (Wilkinson, 1997a). Three of these regions surround the Cys/Asp/His catalytic triad residues, and the other three are of unknown function (D'Andrea and Pellman, 1998). The UBPs vary greatly in size, from 526 amino-acids (DUB1) to 2691 (FAF protein), and have either N- or C-terminal extensions. These are thought to function in substrate determination or cellular localization (D'Andrea and Pellman, 1998). The UBP family is considerably more diverse than the UCHs, demonstrated by the existence of 16 UBPs in the *S. cerevisiae* genome, in comparison with only one UCH (D'Andrea and Pellman, 1998).

Interestingly, depending on the point in the Ub-proteasome pathway a DUB has its effect, it may promote either the degradation or stabilization of a given protein substrate. For example, ubp14, Doa4 and Ap-uch promote proteasome-mediated degradation of substrate proteins,

while FAF and PA700 act to repress proteolysis (D'Andrea and Pellman, 1998). At the moment, the general dogma of DUBs is that they antagonise ubiquitination in all Ub-dependant pathways, and can thus be considered as equally important regulators of ubiquitinated substrates.

1.7. Ubiquitin-like Proteins

Although the number of Ub conjugating and deconjugating enzymes is apparently very large, many are only proposed to be Ub-specific enzymes by consequence of their sequence similarity with characterised enzymes of Ub pathways. Over the past five years some of the uncharacterised enzymes of the Ub attachment and removal families have been shown not to be involved in Ub metabolism as first thought, but actually perform similar roles for novel proteins with sequence similarity to Ub. These new protein modifiers belong to an emerging family of proteins known as the Ub-like proteins (ULPs). As their numbers have increased, and functions characterised the ULPs have been divided into two distinct categories or classes, dependent on whether or not they are thought to play a role as post-translational modifiers. The two groups are known as the Ub-domain proteins (UDPs), and the Ub-like modifiers (UBLs) (Jentsch and Pyrowolakis, 2000).

1.7.1. Ubiquitin domain proteins

The UDPs (Jentsch and Pyrowolakis, 2000), or the type-2 UBPs (Tanaka *et al.*, 1998) (listed in Table 2B), other than their common Ub-like

Table 2. The two classes of Ub-like proteins (ULPs). **A.** Type-1 ULPs, known as the Ub-like modifiers (UBLs) are either known to, or thought to function as modifiers of proteins in a manner analogous to that of Ub. **B.** Type-2 ULPs, known as the Ub domain proteins (UDPs) bear unique Ub-like domains, but do not form conjugates with other proteins. *The length of Ub prior to cleavage varies dependent on the particular primary gene product from which it derives. Tables compiled in part, using information taken from (Jentsch and Pyrowolakis, 2000; Tanaka *et al.*, 1998).

A

Ubiquitin-like Modifiers (UBLs/Type-1 ULPs)	Number of Residues (After Cleavage)	Conjugated to Proteins in a Manner Analogous to Ubiquitination	Sequence Similarity to Ubiquitin	Requires Maturation by C-terminal hydrolase	Contains C-Terminal Gly-Gly Motif	Known to Act as a Protein Modifier
Ubiquitin	(76)*	✓	✓	✓	✓	✓
SUMO-1/SMT3	101 (97)/ 101 (97)	✓	✓	✓	✓	✓
SUMO-2	103 (92)	✓	✓	✓	✓	✓
SUMO-3	96 (93)	✓	✓	✓	✓	✓
NEDD8/Rub1	81 (76)/ 77 (76)	✓	✓	✓	✓	✓
UCRP	165 (157)	✓	✓	✓	✓	✓
FAT10	165	?	✓	✗	✓	?
HUB	73	✓	✓	?	✗	✓
Fau	133 (74)	?	✓	✓	✓	?
An1a, An1b	693 (103)/ 701(103)	?	✓	✓	✓	?
APG12	140	✓	✗	✗	✗	✓
URM1	99	✓	✗	✗	✓	✓

B

Ubiquitin Domain Proteins (UDPs/Type-2 ULPs)	Number of Residues	Position of Ubiquitin-like Domain
RAD23/HHR23A, B	398/363/409	N-(7-63)
DSK2	373	N-(16-71)
PLIC-1	582	N-(~50-110)
PLIC-2/Chap1	638	N-(~50-110)
XDRP1	585	N-(~22-80)
BAG-1L	345	Central-(151-299)
BAG-1S	230	Central-(136-184)
BAT3/Chap-2	1229	N-(42-112)
Scythe	1135	N-(40-110)
Parkin	465	N-(1-72)
UIP28/RBCK1	498	N-(1-72)
UBP6	499	N-(1-65)
Elongin B	118	N-(1-59)
Gdx	157	N-(1-72)

domains, are generally unrelated in sequence to each other. Their Ub-like domains are most commonly found in the N-terminal region of the proteins, and are thought not to be cleaved out or conjugated to other proteins. The biological roles of UDPs are largely unknown. UDPs are thought to play a role in proteasome adaptation. In particular RAD23 and BAG-1 are known to associate with the 26S proteasome in a manner dependant on their Ub-like domains (Uds) (Watkins *et al.*, 1993). RAD23 is involved in yeast nucleotide excision repair and the proteasome protein required for its interaction is RPN10 (S5a), which as discussed earlier is known to recognise Ub moieties and form part of the 19S complex (Hiyama *et al.*, 1999; Schauber *et al.*, 1998). The precise function of this interaction is not fully understood, but a yeast mutant deficient in RAD23 and the functionally overlapping DSK2, is temperature sensitive in growth due to a block in spindle pole body 'centrosome' duplication (Biggins *et al.*, 1996). The growing list of UDPs includes proteins that play active roles in the ubiquitination system. Namely the DUB UBP6, and elongin B which is part of the CBC^{VHL} Ub-ligase (Martinez-Noel *et al.*, 1999; Morett and Bork, 1999). Whether all UDPs associate with the proteasome is currently disputed, but most data are consistent with a model in which UDPs have a role as adaptors that link the ubiquitination machinery to functions of the proteasome.

1.7.2. Ubiquitin-like modifiers

With the exception of APG12, URM1 and FAT10, the UBLs (Jentsch and Pyrowolakis, 2000), or the type-1 ULPs (Tanaka *et al.*, 1998), (listed in Table 2A) are synthesised as precursor forms with short C-terminal extensions, requiring the cleavage of the extrapeptide for maturation. APG12 and URM1 are the only UBLs with no sequence homology to Ub, and owe their classification as UBLs to the Ub-like characteristics of their conjugation pathways (Klionsky and Ohsumi, 1999). UBLs are attached covalently to other proteins by their C-termini which in most cases contain a di-glycine motif capable of ligating to target proteins in a pathway resembling that of Ub modification. Although it should be noted that some members of the group (FAT10, Fau, An1a and An1b) are only included due to their Ub homology and diglycine motifs, and have not yet been shown to be bona-fide modifiers (Jentsch and Pyrowolakis, 2000). Two mammalian proteins Ub cross-reacting protein (UCRP) and FAT10 contain two Ub-like domains fused in tandem (Liu *et al.*, 1999; Loeb and Haas, 1994). Preliminary studies of Ub cross-reacting protein (UCRP) indicate that it is primarily synthesised as a 17kDa precursor which is cleaved to produce a 15kDa protein which can form cellular protein conjugates associated with intermediate filaments (Loeb and Haas, 1994). An interesting member of the UBL family is the modifier HUB (Homologous to Ub). HUB is only 22% identical to Ub and is highly conserved in yeast (Jentsch and Pyrowolakis, 2000). HUB is unique in that it possesses a dityrosine motif at its C-terminus in place of the more common diglycine found in many of the other members. However, little more is known

Figure 3

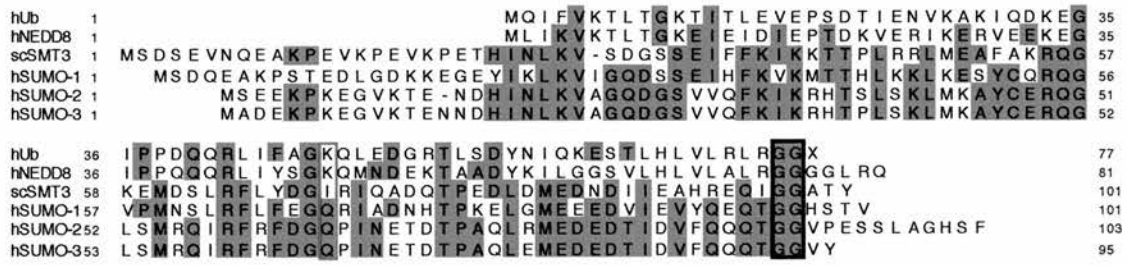


Table 3

	Percentage Primary Sequence Identity After Maturation					
Modifier	hUb	hNEDD8	scSmt3	hSUMO-1	hSUMO-2	hSUMO-3
hUb	100	58	13	18	16	16
hNEDD8	58	100	21	21	18	18
scSmt3	13	21	100	54	50	51
hSUMO-1	18	21	54	100	52	52
hSUMO-2	16	18	50	52	100	96
hSUMO-3	16	18	51	52	96	100

Figure 3 & **Table 3.** Primary sequence alignments and identity comparisons between Ub and the Ub-like modifiers NEDD8, Smt3, SUMO-1, SUMO-2 and SUMO-3. **Figure 3.** Primary sequence alignments of the human proteins Ub (hUb), hNEDD8, hSUMO-1, hSUMO-2, hSUMO-3 and the *S. cerevisiae* homologue of SUMO-1 (scSMT3 also known as Smt3p) (Accession numbers AAA36789, XP007307, CAA67898, CAA67896, CAA67895 and NP_010798 respectively). The diglycine motif is shown boxed in black, and the alignments were based on known structures, although alignments giving higher sequence homology can be made which do not leave N-terminal extensions. Corresponding residues to Lys48 of Ub known to be required for branched Ub conjugation are boxed in red. Identical residues are shown in cyan. **Table 3.** Primary sequence identity comparisons between hUb and the other UBLs compared in the alignment shown in Figure 3. Identity calculations were based on reasonably maximised alignments of the two compared sequences as opposed to the generalised alignment shown in Figure 3.

about UCRP or HUB at present. Another recently identified member of the UBL family is URM1 from yeast, which is unrelated to either Ub or APG12 (Furukawa *et al.*, 2000). Interestingly, the E1 enzyme for the URM1 conjugation pathway, Uba4 is closely related to microbial sulfur-transferases, ThiF and MoeB involved in the synthesis of thiamin and the molybdopterin cofactor (Rajagopalan, 1997; Taylor *et al.*, 1998; Unkles *et al.*, 1999). Sulphur transfer chemistry in both cases involves the addition of a sulfur atom to the C-terminal carboxyl group of a short polypeptide ThiS or MoaD, which both terminate in a diglycine motif. The C-termini of ThiS and MoaD are, like Ub, first activated by the E1-like enzymes through adenylation. The link URM1 provides between mammalian Ub/UBL conjugation, and microbial sulfur-containing enzyme cofactor biosynthesis is evolutionarily important, implying that UBLs may have ancient origins (Hochstrasser, 2000).

1.7.2.1. NEDD8/Rub1 conjugation

Although the available information for the majority of UBLs is limited, two in particular, SUMO-1 and NEDD8 (neuronal precursor cell-expressed developmentally down-regulated protein 8) have been relatively heavily researched. NEDD8 is found in cells as either a free monomer of approximately 6 kDa, or in a series of high molecular weight protein conjugates (Kamitani *et al.*, 1997a). NEDD8 has one of the highest degrees of sequence similarity with Ub of all known ULPs (57% identical), and contains the conserved C-terminal diglycine motif essential for protein

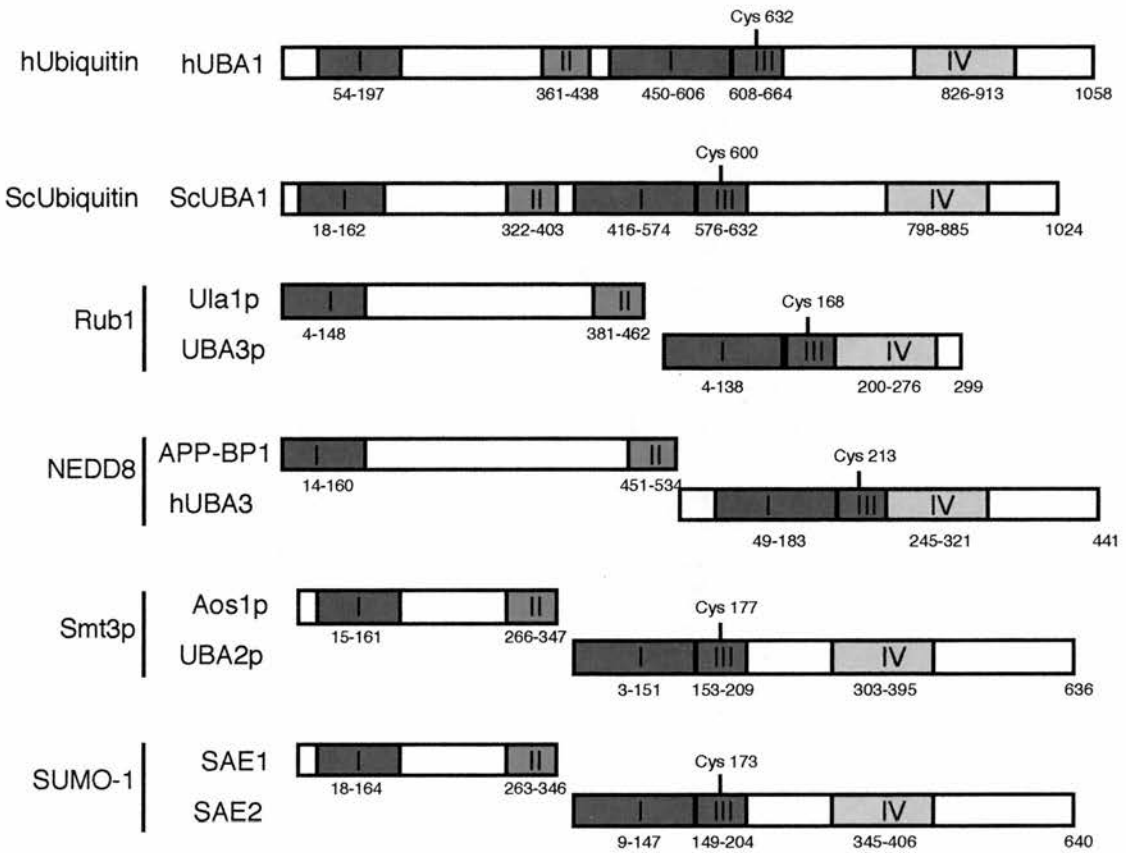
attachment (alignments and identity comparisons shown in Figure 3 and Table 2). Conjugation of NEDD8 and its yeast homologue Rub1 (related to Ub) to cellular proteins follows a pathway analogous to but distinct from Ub conjugation, as shown schematically in Figure 2. NEDD8 conjugation involves a heterodimeric E1 activating enzyme (APP-BP1/hUBA3), and an E2 conjugating enzyme (Ubc12). APP-BP-1 (amyloid precursor protein binding protein 1) as its name suggests, was identified before the discovery of NEDD8, by its interaction with β-amyloid precursors (Chow *et al.*, 1996), while UBA3 and Ubc12 were thought originally to be a Ub-specific enzymes (Liakopoulos *et al.*, 1999; Osaka *et al.*, 1998). Together APP-BP1 and hUBA3, and their yeast homologues ULA1 and UBA3, are known to contain sequence similarity to the N- and C-terminal portions of the monomeric Ub-specific E1 enzymes, thus classifying them as members of an ever-increasing family of proteins with E1-like characteristics (Figure 4). Although Lys 48 is conserved in NEDD8 (see Figure 3), like all UBLs, NEDD8 is thought not to form polyNEDD8 branched forms *in vivo* or *in vitro*. However, recent studies have shown that NEDD8 can be conjugated to Ub via Lys 48 in a branch-like manner and still target the chimerically modified protein to the proteasome (Whitby *et al.*, 1998), suggesting that mechanisms for Ub/NEDD8 hetero-polymer formation may exist. Unlike Ub, the NEDD8 conjugation system is thought to act only on a few substrates, the members of the cullin protein family. In fact, the major substrate for Rub1 in yeast is Cdc53 (known as the yeast cullin), required for G1-S cell cycle progression (Lammer *et al.*, 1998; Liakopoulos *et al.*, 1998). Cullins function to provide scaffolding for large oligomeric E3 Ub-

ligases of the SCF-type (Deshaies, 1999). SCF the first discovered member of the family, as mentioned previously comprises CUL-1, Skp1 and one of a number of F-box proteins. Proteins of this complex also interact with the RING finger protein Roc1 (RBX1), which recruits Ub E2 enzymes such as $UBC3^{CDC34}$ to the ligase complex (Deshaies, 1999). The function of SCF is vital for cells, with substrate specificity for cyclins, CDK inhibitors, β-catenin and IκBα. Human CUL-2 is part of a different ligase complex which also contains the affore-mentioned UDP elongin B, elongin C, Roc1 and one of several substrate selecting SOCS-box proteins that are distinct from F-box proteins (Deshaies, 1999). Although ligase enzymes are not thought to be essential for NEDD8 conjugation, it has been suggested that *in vivo* the efficiency of NEDD8 conjugation may be increased by E3-like proteins. One of these factors is the tumor suppressor pVHL (von Hippel-Lindau protein), which when associated in the complex forms the CBC^{VHL} (Wada *et al.*, 1999). NEDD8 modified cullins are found exclusively within the assembled Ub-ligase complex (Kamura *et al.*, 1999; Lammer *et al.*, 1998; Liakopoulos *et al.*, 1999). However, expression of a tumorigenic variant of pVHL lacking the CBC interaction domain fails to induce the NEDD8 modification of cullins which occurs in the wild-type (WT) system (Liakopoulos *et al.*, 1999). This suggests that the CBC^{VHL} E3 ligase has to fully assemble before the modification occurs, indicating that the complex not only mediates ubiquitination, but also autocatalyses CUL-2 modification by NEDD8. NEDD8 conjugation is thought not to be essential for the complex activity as yeast Rub1-null mutants are viable (Liakopoulos *et al.*, 1999; Osaka *et al.*, 1998). Further studies have, however,

implicated NEDD8 as a stabilising factor in the complex. Due to the limited diversity of NEDD8/Rub1 protein substrates (Cdc53, Cul-1, Cul-2 Cul-3 Cul-4A, Cul-4B and Cul-5), consensus motifs of modification sites may be of little value in the prediction of novel substrates. At present the motif IxxIM**K**xbb (where x is any residue **K** is the target lysine, and b represents a basic residue) appears to be conjugation consensus, but if or when other non-cullin substrates are identified, this may need revision.

The distinct protease activities responsible for the maturation of NEDD8 by removal of the inhibitory C-terminal peptide, and the removal of NEDD8 from conjugated proteins have been identified as previously considered Ub-specific enzymes (Gong *et al.*, 2000a; Wada *et al.*, 1998). However, both the C-terminal hydrolase and isopeptidase activities, known as UCH-L3 (Wada *et al.*, 1998) and USP21 (Gong *et al.*, 2000a) respectively in mammals, appear not to be entirely specific to NEDD8 alone, as they have dual specificity for Ub and NEDD8, but not for other UBLs. Little is known about the structure of the isopeptidase USP21 other than it has a calculated molecular mass of 43 kDa, and that it contains the highly conserved Cys, Asp and His domains found in all Ub isopeptidases. Recently, the structure of the NEDD8/Ub C-terminal hydrolase UCH-L3 has been resolved by crystallography, and was shown to resemble the papain-like cysteine proteases, particularly cathepsin B (Johnston *et al.*, 1997). Unfortunately, at present, little is known about their physiological roles or regulation *in vivo* and as such the regulatory capacity of their activities is poorly understood. In general, the current findings imply that

A


hUbiquitin
hUBA1
Cys 632
I
II
I
III
IV
54-197
361-438
450-606
608-664
826-913
1058
ScUbiquitin
ScUBA1
Cys 600
I
II
I
III
IV
18-162
322-403
416-574
576-632
798-885
1024
Rub1
Ula1p
UBA3p
Cys 168
I
II
I
III
IV
4-148
381-462
4-138
200-276
299
NEDD8
APP-BP1
hUBA3
Cys 213
I
II
I
III
IV
14-160
451-534
49-183
245-321
441
Smt3p
Aos1p
UBA2p
Cys 177
I
II
I
III
IV
15-161
266-347
3-151
153-209
303-395
636
SUMO-1
SAE1
SAE2
Cys 173
I
II
I
III
IV
18-164
263-346
9-147
149-204
345-406
640

B


hUbc7 1 -MAGTALKRLMAVYKQLTLNPPEGIVAGPMNE----ENFFEWEAL 40
scUbc7 1 -MSKTAQKRLLKELQQLIKDSPPGIVAGPKSE----NNIFIWDCL 40
hUbc12 1 MIKLFSLKQQKKEEESAGGTKGSSKKASAAQLRIQKDINELN--LPKTCDISFSDP----DDLLNFKLV 63
scUbc12 1 MLKLRQL--QKKKQKENENSSSIQPNLSAARIRIKRDLDSLD-LPPTVTLNVLTSPDSADRSQSPKLEV 66
hUbc9 1 -MSGIALSRIAQERKAWRKDHPFGFVAVPTKNPDGTMNLMNWECA 44
scUbc9 1 -MSSLCLQRLQEERKKWRKDHPFGFYAKPVKKADGSMDLQKWEAG 44
hUbc7 41 IMGPEDTCFEFGVFPAILSFPLDYPLSPPKMRFTCEMFHPNIYPDGRVCISILHAPGDDPHGLREQPER 109
scUbc7 41 IQGPPDTPYADGVFNAKLEFPKDYPLSPPKLTFTPSILHPNIYPNGEVCISILHSPGDDPNMYELAEER 109
hUbc12 64 I-CPDEGFYKSGKFVFSFKVGQGYPHDPPKVKCETMVYHPNIDLEGNVCLNILRE----------D 118
scUbc12 67 IVRPDEGYYNYGSINFNLDFNEVYPIEPPKVVCLKKIFHPNIDLKGNVCLNILRE----------D 122
hUbc9 45 IPGKKGTPWEGGLIKLRMLIKDDYPSSPPKCKFEPPLFHPNVYPSGTVCLSILEEDK----------D 102
scUbc9 45 IPGKEGTNWAGGVYPITVEYPNEYPSKPPKVKFPAGFYHPNVYPSGTICLSILNEDQ----------D 102
hUbc7 110 WSPVQSVEKILLSVMSMLSEPNIESGANIDACILWRDNRPEFERQVKLSILKSLGF 165
scUbc7 110 WSPVQSVEKILLSVMSMLSEPNIESGANIDACILWRDNRPEFERQVKLSILKSLGF 165
hUbc12 119 WKPVLTINSIIYGLQYLFLEPNPEDPLNKEAAEVLQNNRRLFEQNVQRSMRGGYIGSTYFERCLK 183
scUbc12 123 WSPALDLQSIITGLLFLFLEPNPNDPLNKDAAKLLCEGEKEFAEAVRLTMSGGSIEHVKYDNIVSP 188
hUbc9 103 WRPAITIKQILLGIQELLNEPNIQDPAQAEAYTIYCQNRVEYEKRVRAQAKKFAPS 158
scUbc9 103 WRPAITLKQIVLGVQDLLDSPNPNSPAQEPAWRSFSRNKAEYDKKVLLQAKQYSK 157

in comparison with Ub itself, NEDD8 is a relatively selective modification which is affiliated with the Ub-proteasome system.

1.7.2.2. SUMO-1 conjugation

The small Ub like modifier (SUMO-1) was originally identified in vertebrate systems independently by a number of different groups using various two-hybrid screens or biochemical analyses, and subsequently given a number of different titles; sentrin 1 (Okura *et al.*, 1996), UBL1 (Shen *et al.*, 1996), PIC1 (Boddy *et al.*, 1996), GMP1 (Matunis *et al.*, 1996), and SMT3c (Lapenta *et al.*, 1997). SUMO-1 has a much lower degree of similarity with Ub than NEDD8 (18% identical) (see Figure 3 and Table 3), but is thought to have a much higher substrate diversity than NEDD8, with a more subtle and complex range of cellular significances. Amongst numerous other species, the budding and fission yeast and insect homologues of vertebrate SUMO-1 have been identified known as Pmt3p (Tanaka *et al.*, 1999), Smt3p (Johnson *et al.*, 1997; Meluh and Koshland, 1995) and DmSmt3 (Huang *et al.*, 1998) respectively. Two other members of the SUMO family have been identified, named SUMO-2 (also known as SMT3a and sentrin-3) (Lapenta *et al.*, 1997) and SUMO-3 (also known as SMT3b, sentrin-2 and originally HSMT3) (Mannen *et al.*, 1996). SUMO-2 and SUMO-3 are often referred to as "twins" as they are 97% identical to each other but only 47% identical to SUMO-1 (in the Ub homology domain) (see Figure 3 and Table 3 for comparisons). SUMO-2 and SUMO-3 have not been investigated as thoroughly as SUMO-1, and as such little

data exists regarding their cellular role, although preliminary investigations indicate they perform functionally distinct roles from SUMO-1 (Saitoh and Hinchey, 2000). Like Ub and NEDD8 all members of the SUMO family are primarily synthesised as inactive precursors with distinct C-terminal peptide extensions which are removed by specific protease activities to expose a diglycine motif (Gly-Gly), of which the C-terminal Gly is involved in their attachment to substrates (schematic of activation and conjugation shown in Figure 2).

1.7.2.2.1. SUMO-1 conjugation enzymes

The enzymes involved in the conjugation of SUMO-1 in vertebrate systems (Desterro *et al.*, 1999; Desterro *et al.*, 1997; Gong *et al.*, 1997; Gong *et al.*, 1999; Okuma *et al.*, 1999), Smt3p in budding yeast (Johnson and Blobel, 1997; Johnson *et al.*, 1997) and DmSmt3 in *Drosophila* (Bhaskar *et al.*, 2000) have been isolated and characterised to varying degrees. The SUMO-1/Smt3p conjugation system(s) are generally described as following a pathway similar to, yet distinct from, Ub conjugation, although a number of important differences exist between the two (Figure 2). SUMO-1/Smt3p conjugation to target proteins, appears to require the activities of only two enzymes, specific E1 and E2 enzymes *in vitro*, with little or no evidence for the requirement of an E3 ligating enzyme (Desterro *et al.*, 1997; Okuma *et al.*, 1999). Like APP-BP1/UBA3 for NEDD8 and unlike UBA1 for Ub, SUMO-1 and Smt3p are activated in the first step by a heterodimeric E1 activating enzyme. These proteins are known to have sequence homology

with the monomeric Ub E1, UBA1 (Desterro *et al.*, 1999) (see alignment in Figure 4). The four conserved regions found in the monomeric UBA1 spread over the individual proteins of the dimeric SAE1/SAE2 (for SUMO-1) and Aos1p/Uba2p (for Smt3p). The role of the E1 enzyme heterodimers in SUMO-1 conjugation, is thought to be essentially identical to that of UBA1 for Ub. Although not as thoroughly characterised as the Ub E1, it is known that SAE1/SAE2 forms an ATP-dependent thioester bond with the C-terminal Gly residue (Gly97) of SUMO-1 via the putative Cys 173 of SAE2 (Desterro *et al.*, 1999; Okuma *et al.*, 1999). Both monomers of the dimeric E1s are known to be required for activation of SUMO-1/Smt3p and transfer to the E2 SUMO-1-conjugating enzyme. In the budding yeast system the E2 activity is known to be Ubc9p (Johnson and Blobel, 1997), with hUbc9 being the human homologue (Desterro *et al.*, 1997; Gong *et al.*, 1997; Schwarz *et al.*, 1998), again previously thought to be Ub-specific enzymes (generally referred to collectively as Ubc9). The transesterification of activated SUMO-1 from SAE2 to Ubc9 occurs on Cys93 of the E2 protein. In the final step, Ubc9 recognises substrates itself, and subsequently catalyses the formation of an ε-amino bond with lysine residues on the target protein. The necessity of an E3 for the SUMO-1 conjugation system *in vivo* is still a matter of debate. Interestingly, unlike ubiquitinated lysine residues where no single general consensus exists, targets for SUMO-1 modification are almost always found within a **ΨK**xE consensus motif (where Ψ represents any large hydrophobic residue and x any residue) (Rodriguez *et al.*, 1999). The fact that a Ub consensus motif for

all E2s does not exist, suggesting that specificity for substrate is largely carried by the E3s, lends weight to the argument that E3s are unlikely to be absolutely required for *in vivo* SUMO-1 conjugation. However, it is not unfeasible that *in vivo* proteins exist that increase the efficiency of, but are not essential for SUMO-1 conjugation to particular targets, in an E3 like manner.

1.7.2.2.2. Substrates of SUMO-1 conjugation

A large number of proteins have been shown to interact with Ubc9 in yeast two-hybrid assays and although some have been consequently shown to be SUMO-1-conjugation substrates, many have not (see Yeh *et al.*, 2000 for current list). It is possible that many of them are not modified, and may in fact be regulated by (or regulate) Ubc9 via a mechanism independent of SUMO-1 conjugation (Huggins *et al.*, 1999; Long and Griffith, 2000; Poukka *et al.*, 1999; Xu *et al.*, 2000). The number of proteins modified by SUMO-1 is constantly increasing (see Table 4 for current list). Many of those identified recently as substrates (such as the Werner's disease protein Wrn (Kawabe *et al.*, 2000), the p53 homologue p73α (Minty *et al.*, 2000), the DNA topology isomereases topoisomerase I and II (Mao *et al.*, 2000a; Mao *et al.*, 2000b), and the glucose transport proteins GLUT1 and GLUT4 (Giorgino *et al.*, 2000)) have not been characterised thoroughly, and thus any functional relevance is mostly based on speculation. The majority of the SUMO-1 conjugated proteins are localised in the nucleus or at the nuclear envelope, although cytoplasmic and nucleocytoplasmic

Species System	SUMO-1-Modified Protein	Cellular locality	Effects of Modification on Substrate and/or Cell	References
Vertebrates	PML	Nucleus	SUMO-1-PML found in PML nuclear bodies, and is required for Daxx recruitment.	(Duprez et al., 1999; Ishov et al., 1999; Muller et al., 1998a; Sternsdorf et al., 1997)
	Sp100	Nucleus	SUMO-1-SP100 is nuclear localised	(Sternsdorf et al., 1997)
	p53	Shuttling	SUMO-1-p53 has increased transactivation	(Gostissa et al., 1999; Rodriguez et al., 1999)
	Mdm2	Cytoplasm	SUMO-1-Mdm2 is resistant to auto-ubiquitination, thus stabilised. Results in decreased transactivation by p53.	(Buschmann et al., 2000)
	IκBα	Shuttling	SUMO-1- IκBα is resistant to Ub-mediated degradation. Inhibits NF-kB-dependent gene transcription.	(Desterro et al., 1998)
	RanGAP1	Nuclear envelope	SUMO-1-RanGAP1 is targeted to the nuclear pore complex.	(Mahajan et al., 1997; Mahajan et al., 1998)
	Ran BP2	Nuclear envelope	Unknown	(Saitoh et al., 1998)
	HIPK2	Nucleus	SUMO-1-HIPK2 required for formation of nuclear speckles distinct from PML nuclear bodies.	(Kim et al., 1999)
	GLUT1	Cell Membrane	Involved in regulation of insulin-mediated glucose uptake.	(Giorgino et al., 2000)
	GLUT4	Cell Membrane	Involved in regulation of insulin-mediated glucose uptake.	(Giorgino et al., 2000)
	TEL	Nucleus	SUMO-1-TEL found in TEL nuclear bodies.	(Chakrabarti et al., 2000)
	p73α	Shuttling	Affects subcellular localisation and stability.*	(Minty et al., 2000)
	DNA topo-isomerase I	Nucleus	Involved in cell response to TOP-1-mediated DNA damage.	(Mao et al., 2000b)
	DNA topo-isomerase II	Nucleus	Involved in cell response to TOP-2-mediated DNA damage.	(Mao et al., 2000a)
	Wm	Nucleus	Unknown	(Kawabe et al., 2000)
Yeast	Cdc3	Nucleus	Smt3-Cdc3 is involved in regulation of septin-ring dynamics during cell cycle.	(Johnson and Blobel, 1999; Takahashi et al., 1999)
	Cdc11	Nucleus	Smt3-Cdc11 is involved in regulation of septin-ring dynamics during cell cycle.	(Johnson and Blobel, 1999)
	Shs1/Sep7	Nucleus	Smt3-Sep7 is involved in regulation of septin-ring dynamics during cell cycle.	(Johnson and Blobel, 1999)
Drosophila	Tramtrack 69	Nucleus	Unknown	(Lehembre et al., 2000)
	Drosophila Dorsal	Nucleus	Nuclear targeting.*	(Bhaskar et al., 2000)
	CaM Kinase II	Cytoplasm	Neuronal differentiation.*	(Long and Griffith, 2000)
Viral (Using host cell machinery)	CMV IE1	Nucleus	Regulation of host PML NB structure.*	(Muller and Dejean, 1999)
	BPV E1	Nucleus	Nuclear accumulation.	(Rangasamy and Wilson, 2000; Rangasamy et al., 2000)
	CMV IE2-p86	Nucleus	Transactivation	(Hofmann et al., 2000)
	HPV 1a	Nucleus	Unknown	(Rangasamy and Wilson, 2000)
	HPV 18 E2	Nucleus	Unknown	(Rangasamy and Wilson, 2000)

Table 4. Summary of known substrates for SUMO-1 modification to date in vertebrates, yeast, drosophila, and viruses. * Denotes that conclusive data has not yet been presented.

shuttling examples have also been demonstrated. Targets for SUMO-1 modification have been identified in vertebrates, yeast and *Drosophila* and viral systems in host cells (shown in Table 4.). It is clear that although SUMO-1 appears to have a significant role in nuclear and subnuclear targeting, novel cellular functions of the modification pathway are constantly being discovered. As such, due to the diverse substrate-specific affects, the relevance of SUMO-1 modification cannot be broadly generalised.

1.7.2.2.1.1. SUMO-1 substrates found in sub-nuclear structures

Whether by coincidence, or a genuine representation of the profile of substrates, many SUMO-1-conjugated proteins are known to associate in specific complexes forming sub-nuclear structures. One of the most heavily researched examples of such a protein is the promyelocytic leukaemia protein (PML). Acute promyelocytic leukaemia (APL) is characterised by a specific t(15;17) chromosomal translocation that fuses the genes encoding PML and the retinoic acid receptor α (RARα), generating a PML-RARα chimera (de The *et al.*, 1991). The biological function of PML is poorly understood, although it is known to have tumor suppressor activity, and localises in distinct subnuclear domains known as PML nuclear bodies (NBs), also known as ND10s or PODs, that are tightly associated with the nuclear matrix (Dyck *et al.*, 1994). These structures appear to be disrupted in APL cells by the PML-RARα protein, but can be rescued by treatment with trivalent antimonyls such as arsenic (As_2O_3) and

antimony tetroxide (Sb_2O_3) which triggers the degradation of PML-RARα (Muller *et al.*, 1998b; Shao *et al.*, 1998). Both PML and PML-RARα have been shown to be SUMO-1 modified (Boddy *et al.*, 1996; Kamitani *et al.*, 1998c; Muller *et al.*, 1998a; Sternsdorf *et al.*, 1997), which has subsequently been shown to be stimulated by trivalent antimonyls. SUMO-1-modified forms of wild-type (WT) PML are metabolically stable and localise in NBs, whereas SUMO-1 modified PML-RARα results in a SUMO-1 specific degradation (Duprez *et al.*, 1999), explaining the observed consequences of APL-cell treatment. PML is SUMO-1 modified on lysines 65, 160 and 490 (Kamitani *et al.*, 1998b) in the nucleus, and it is possible that modification may be a targeting mechanism for the NBs (Duprez *et al.*, 1999). Furthermore, SUMO-1 modified PML is known to be essential for the recruitment of another protein, Daxx, to the PML NBs (Ishov *et al.*, 1999). Another component of NBs, Sp100, thought to be involved in chromatin dynamics (Seeler and Dejean, 1999), is known to be modified by SUMO-1. Sp100 modification occurs within a domain which is known to be involved in homodimerisation and NB-targeting (Sternsdorf *et al.*, 1999). Interestingly, reports of the disruption of NBs in response to infection with DNA viruses such as herpes simplex virus (HSV) and cytomegalovirus (CMV), have shown that both viruses code for proteins which disrupt the SUMO-1 modification of PML and Sp100 (Muller and Dejean, 1999). Furthermore, CMV protein IE1 which mediates this disruption, is itself SUMO-1 modified *in vivo* (Muller and Dejean, 1999), implicating the SUMO-1 system may also be important for viral life cycles.

Other nuclear complex and structure forming proteins have recently been identified as SUMO-1 substrates. The transcription factor/tumor suppressor TEL, a product of the E-26 transforming specific-related gene, is one such protein (Chakrabarti *et al.*, 2000). TEL was originally shown to bind Ubc9 via its helix-loop-helix (HLH) motif (Chakrabarti *et al.*, 1999), and in fact, Lys99 within this domain is known to be the target for SUMO-1 modification. The modification is essential for the localisation of TEL into cell-cycle specific nuclear speckles called TEL bodies (Chakrabarti *et al.*, 2000), the cellular significance of which, has yet to be elucidated. Consistent with the trend, the homeodomain-interacting protein kinase (HIPK2), has recently been shown to be targeted to nuclear bodies distinct from PML NBs, when modified by SUMO-1. HIPK2 is a member of the family of nuclear protein kinases that act as corepressors for homeodomain transcription factors. Like Sp100, the modification site is thought to exist within a domain known to be essential for sub-nuclear structure association (Kim *et al.*, 1999). All of these studies lend weight to the theory that SUMO-1 may play a significant role in the general assembly and/or activity of complex nuclear proteins which form distinct subnuclear structures.

1.7.2.2.1.2. Other SUMO-1 substrates

The first example of a protein covalently linked to SUMO-1 was the mammalian guanosine triphosphate (GTP)ase-activating protein RanGAP1 (Bischoff *et al.*, 1995; Mahajan *et al.*, 1997; Matunis *et al.*, 1996).

Asymmetric distribution of the small GTPase Ran and associating proteins between the nucleus and the cytoplasm is thought to be responsible for the vectoral nature of nucleocytoplasmic transport (Stochaj *et al.*, 1998). In eukaryotes exchanges between the nucleus and the cytoplasm transit the nuclear envelope (NE) through the massive protein-composed channels, nuclear pore complexes (NPCs), that are embedded in the NE (Nigg, 1997). Nuclear protein import targets proteins that contain nuclear localisation signals (NLSs) and is mediated by soluble transport factors as well as proteins of the NPC, termed nucleoporins (Nigg, 1997). Ran is an abundant protein, which is located mostly in the nucleus. Mutation of Ran is known to directly affect the nuclear import of proteins with basic NLSs (Stochaj *et al.*, 1998). Its GTPase activity is stimulated by the regulatory proteins, RanGAP, and nuclear guanidine nucleotide exchange factor (GEF). RanGAP1 is known to be essential for the transport of proteins into the nucleus (Melchior *et al.*, 1993; Moore and Blobel, 1993). RanGAP1 is localised predominantly in the NE, where it forms a stable complex with the NPC protein Ran binding protein 2 (RanBP2) (Mahajan *et al.*, 1997; Matunis *et al.*, 1996). Investigations into this process revealed that all NPC associated RanGAP1 is SUMO-1 modified at a single lysine (Lys526) (Mahajan *et al.*, 1998), and the association of RanGAP1 with RanBP2 is in fact SUMO-1 modification-dependent (Matunis *et al.*, 1998), as unmodified RanGAP1 is exclusively cytoplasmic. Additionally, RanBP2 is also known to be SUMO-1 modified, although the significance is not yet understood (Saitoh *et al.*, 1998). The reversible nature of SUMO-1 modification has thus lead to the speculation that the dynamic SUMO-1 modification state

of NPC proteins may be a key controlling factor over nuclear import of basic NLS containing proteins.

As mentioned, although the majority of SUMO-1 modified proteins are localised in the nucleus, or associated with the NE, the shuttling protein IκBα is known to be a target for SUMO-1 modification (Desterro *et al.*, 1998). In unstimulated cells the transcription factor NF-kB is retained in the cytoplasm in an inactive state by IκB inhibitor proteins. Exposure of cells to a wide variety of stimuli results in the release of transcription factor from IκB, allowing the active (DNA binding) form of NFκB to translocate into the nucleus where it binds to its recognition sites in the upstream regions of a variety of genes (Hay *et al.*, 1999). Several IκBs have been described, including IκBβ, IκBε, and Bcl-3, but IκBα has been the most thoroughly investigated NFκB inhibitor (Hay *et al.*, 1999). Following signal induction IκBα is rapidly phosphorylated followed by ubiquitination on lysines 21 and 22 (and secondarily lysines 38 and 47), which target it for the Ub-proteasome degradation pathway. IκBα is modified by SUMO-1 (in a phosphorylation independent manner), and is known to be a stabilising signal, rendering the protein resistant to Ub-mediated degradation (Desterro *et al.*, 1998). This resistance is incurred in part by the fact that the target lysine for SUMO-1 modification is Lys21, which protects that residue from Ub-conjugation. Although the ability of SUMO-1 modified IκBα to be Ub-conjugated on the other lysines has not been investigated, singly modified IκBα is resistant to proteasomal degradation (Desterro *et al.*,

1998). This introduces the hypothesis that, at least in this case, ubiquitination and SUMO-1 conjugation are antagonistic, and that SUMO-1 may play a general stabilising role for some proteins degraded by the Ub-proteasome pathway.

The highly regulated, multiply-modified protein p53 has recently been identified as a substrate of SUMO-1 modification (Gostissa *et al.*, 1999; Rodriguez *et al.*, 1999). In normal cells p53 is rapidly turned over via the Ub-proteasome pathway mediated by the p53-Ub E3 Mdm2 (Maki *et al.*, 1996). However, upon stimulation its half-life is significantly increased, which is thought to be majorly responsible for the increased transcription of the p53 dependent cellular genes. It has been shown that p53 is modified by SUMO-1 on a single lysine (386), and thus, unlike IκBα does not compete for the same site as ubiquitination (Rodriguez *et al.*, 1999). This modification is known to be stimulated by UV light and enhances p53-dependant transactivation (Gostissa *et al.*, 1999; Rodriguez *et al.*, 1999). Interestingly, evidence has been presented suggesting that Mdm2 is itself SUMO-1 modified (on Lys446) within the RING finger domain, which is proposed to abrogate its self-ubiquitination activity and subsequently stabilises the protein. Contrary to p53 itself, SUMO-1 conjugation of Mdm2 is thought to be reduced in response to DNA damage, and thus the turnover of Mdm2 and p53 are increased and reduced respectively (Buschmann *et al.*, 2000). Together the dynamic SUMO-1 modification states of p53 and its Ub-ligase Mdm2 may work together to promote p53 stability.

Although the transcription factor c-Jun was shown to interact with Ubc9 over three years ago (Gottlicher *et al.*, 1996), it hasn't been until recently that the SUMO-1 modification of c-Jun has been conclusively proven (Muller *et al.*, 2000). Like p53 and IκBα, c-Jun is a known target for ubiquitination, and similarly it is only SUMO-1 conjugated on a single lysine residue (Lys229). The phosphorylation of c-Jun which is known to inhibit Ub-modification, also inhibits SUMO-1 conjugation. Studies using K229R mutants of c-Jun revealed that SUMO-1 conjugation appears to negatively regulate its activity, although the process has yet to be characterised fully.

1.7.2.3. SUMO-1-specific proteases

The first UBL-specific processing enzyme was published in early 1999 (Li and Hochstrasser, 1999). Ulp1 from yeast was shown to both process Smt3 to its mature form, and specifically cleave proteins from SUMO-1 and Smt3, but not Ub. Interestingly although SUMO-1, NEDD8 and Ub, and their respective E1 and E2 enzymes share significant sequence identity, the 621 residue, 72 kDa Ulp1 is unrelated to any known deubiquitinating enzymes. It does however show distant similarity to certain viral proteases such as the adenovirus L3 protease. Its cellular significance was shown using Ulp1 deficient yeast, indicating it to be essential for G2-M cell-cycle progression (Li and Hochstrasser, 1999), while two-hybrid screening indicated that this may be through its association with nucloeporins (Takahashi *et al.*, 2000). A second member of the yeast

SMT3-specific isopeptidase family was later identified as Ulp2, a product of the SMT4 gene (Li and Hochstrasser, 2000). Studies of Ulp2 deficient yeast suggest that compromised deconjugation activity feedsback on the Smt3 conjugation mechanisms in a negative manner, allowing partial recovery of cell function. The 643 residue human homologue of Ulp1, SENP1 was later cloned and shown to have functional similarity to its yeast counterpart (Gong *et al.*, 2000b). Interestingly it was shown to have specificity for different SUMO-1 conjugated substrates *in vivo* although this is probably an artefact of its cellular localisation as opposed to its specificity, as no substrate specificity was demonstrated *in vitro*. A mammalian SUMO-1-specific enzyme, thought to be one of a number of SUMO-1 proteases with both hydrolase and isopeptidase activity, was identified later by purification from bovine brain (Suzuki *et al.*, 1999). Interestingly it is less than half the size of the yeast Ulp1, being only 30 kDa in mass, although further cloning and characterisation have not yet been published. This bovine protease is also much smaller than the human SUMO-1-specific protease (SUSP1) cloned from human brain, which is approximately 126 kDa in mass (Kim *et al.*, 2000). This protein appears only to have C-terminal hydrolase activity, but is highly specific for SUMO-1, as it is inactive towards Smt3 fusion proteins. Other than the highly conserved His/Asp/Cys catalytic triad found in cysteine proteases, SUSP1 shows no amino-acid sequence homology to any known deubiquitinating enzymes (Kim *et al.*, 2000). Immunofluorescence revealed that SUSP1 is almost entirely cytoplasmic in localisation, indicating this to be its primary region of activity. Six further members of the SENP family of SUMO-1

specific proteases have been identified, and are known to vary in size from 238 (SENP6) to 1112 (SENP7) residues (Yeh *et al.*, 2000). All SENPs have two regions of sequence similarity; a short N-terminal domain of unknown significance, and a conserved C-terminal sequence of ~200 residues which contains the catalytic domain. The function of the varying length inserts between the two domains have not been investigated, but it is likely that they play an analogous role to the inserts in the UBPs with regard to substrate selectivity and sub-cellular localisation. Unfortunately, recombinant SUMO-1 proteases appear to be largely inactive, significantly hindering protein characterisation studies (J.M. Desterro personal comm.) thus limiting further progress in the field.

Until recently investigations have concentrated on the identification and characterisation of the substrates of, and the enzymes responsible for SUMO-1/Smt3 maturation, conjugation and deconjugation, but little is known about the control mechanisms of these processes. Bearing in mind the importance of these for the Ub system, their identification will likely carry a significant degree of gravity.

1.8. Aims of the Project.

The objective of the work described here was to structurally and functionally characterise the mechanisms involved in the conjugation and deconjugation of SUMO-1 to target proteins. This was to be exercised primarily by the solution of the 3D structure of SUMO-1. The functional significance of C-terminal residues proximal to the reactive glycine residue

97, which are of significantly different charge to corresponding residues in Ub, were to be investigated using site-directed mutagenesis and functional analysis. Furthermore, using site-directed mutagenesis, the significance of charged groups found within a variant loop spatially proximal to the active site cysteine 93 residue in Ubc9 were to be investigated for their functional relevance. Finally the work aimed to clarify the identity of the conjugating enzymes responsible for the attachment of SUMO-2 and SUMO-3 to particular substrates, in an effort to determine the mechanisms responsible for the altered substrate specificity noted between themselves and SUMO-1 *in vivo*.

2. MATERIALS AND METHODS

2.1. Materials

All materials not produced in house were purchased from Sigma unless otherwise stated.

2.2. Antibodies

HA-SUMO-1, HA-SUMO-2 and HA-SUMO-3 were detected in Western-blot experiments using mAb 12CA5 (at 1:5000 dilution), which recognises YPYDVPDYA from influenza HA, obtained from BAbCO.

Ubc9 was detected using an affinity purified sheep pAB (in house) at 1:1000 dilution.

SUMO-1 was detected using the mouse anti-GMP1 mAB (Zymed), and SUMO-2 and SUMO-3 were detected using a rabbit anti SUMO-2/-3 pAB (a gift from H.Saitoh Picower Institute, New York) at 1:1000 dilution (Saitoh and Hinchey, 2000).

Sheep anti-mouse (Amersham), donkey anti-sheep (Jackson ImmunoResearch) and donkey anti-rabbit (Amersham) horseradish-peroxidase conjugated IgGs were used to detect the primary antibodies at 1:2500 dilution in Western-blot experiments.

HA-SUMO-1 proteins were detected in immunofluoresence experiments using pAb PRB-101P (1:200 dilution) which also recognises YPYDVPDYA from influenza HA, and was obtained from BAbCO.

PML was detected in immunofluorescence experiments using mAb 5E10 (1:10 dilution) obtained from R. van Driel (Kamitani *et al.*, 1998b).

Secondary antibodies for fluorescence experiments, goat anti-mouse FITC-conjugate, and goat anti-rabbit Texas-Red-conjugate (Harlan), were used at 1:200 dilution.

2.3. Bacterial Strains

E. coli DH5α (genotype: ϕ80d*lac*ZΔM15, *rec* A1, *end* A1, *gyr* A96, *thi*-1, *hsd* R17 (r_k-, m_k+), *sup* E44, *rel* A1, *deo*R, D(*lac*ZYA-*arg*F) U169) was used for routine DNA preparation. *E. coli* B834 (F-, *ompT*, *hsd*S_B, (r_B-, m_B-) *dcm*, *gal*) was used for all protein expression unless otherwise stated. Bacteria were grown in Luria-Bertani broth (LB) supplemented with antibiotics unless otherwise stated. The SUMO-1 mutant T95R in the pET-3a vector was expressed in *E. coli* BL21 DE3 (F-, *omp*T *hsd*S_B (r_B-, m_B-) *gal dcm* (DE3)). For NMR analysis, $^{15}N/^{13}C$ double labelled SUMO-1 was expressed in *E. coli* BL21 (F-, *omp*T *hsd*S_B (r_B-, m_B-) *gal dcm*), grown in minimal medium (as explained in section 2.6).

2.4. Plasmids and Expression Vectors

pGEX-2T was purchased from Pharmacia, the transient mammalian expression vector pcDNA-3 was obtained from Invitrogen Corporation, and pET-3a was purchased from Novogen. The vectors HA-tagged WT-SUMO-$1_{(1-97)}$-pcDNA-3 and Ubc9-pGEX-2T were gifts from Joana Desterro (University of St. Andrews).

2.4.1. DNA preparation

All DNA preparations (minipreps, maxipreps and gel extractions) used for cloning, analysis and transfections were prepared using Quiagen kits as directed by the manufacturer. DNA restriction enzymes were obtained from New England Biolabs (NEB) and Promega. Vent DNA polymerase used in polymerase chain reactions (PCR) was obtained from NEB. Quantity and quality of purified DNA was calculated by spectrophotometric analysis at wavelengths of 260nm and 280nm.

2.4.2. cDNA cloning

2.4.2.1. Preparation of electrocompetant bacteria

A single bacterial colony was picked from an L-B agar plate and used to innoculate 10ml of L-B with no antibiotic selection. The culture was incubated for approximately 12h at 37°C with agitation. This culture was used to innoculate 500ml of L-broth which was subsequently incubated at 25°C with agitation until the culture reached an OD_{600} of between 0.4 and 0.7. The culture was then chilled on ice for 1h before centrifugation at 2500g, at 4°C for 15 mins. The resultant bacterial pellet was resuspended in 250mls of sterile 1mM HEPES pH7.0 chilled to 4°C. This resuspension was centrifuged at 2500g, at 4°C for 15 mins. Resuspension in HEPES buffer and centrifugation was repeated and the pellet of bacteria resuspended in 4 volumes of 10% glycerol solution. Aliquots of cells were then snap frozen in liquid nitrogen and stored at -70°C.

2.4.2.2. Transformation of electrocompetant bacteria

Ethanol precipitated DNA ligations resuspended in 5µl of dH_2O, or 100ng of whole plasmid DNA was mixed on ice with 40µl of electrocompetant bacterial suspension (see above). The DNA/cell suspension was transferred to a 2mm electroporation couvette (Flowgen) previously chilled on ice. Cells were electroporated at 2500V, 201mF for 5msec (Equibio Easyject plus) before addition of 1ml of L-B, 20mM glucose solution followed by inubation at 37°C for 1h. The cell suspension was centrifuged briefly at 14000g and 900µl of supernatant removed. Pelleted cells were resuspended in the remaining 100µl of medium and spread on an L-agar plate containing 100µg.ml^{-1} ampicillin before incubation at 37°C. Tranformed bacteria grew into single colonies overnight.

2.4.2.3. Preparation of thermocompetant bacteria

Bacteria were prepared as described above for electrocompetant cells to the point of washing. For the preparation of thermocompetant bacteria, cells were washed with 10 pellet volumes of sterile 50mM $CaCl_2$, 20mM $MgSO_4$ at 4°C. The cell suspension was centrifuged at 2500g for 15 mins and the pellet resuspended in 4 pellet volumes of sterile 50mM $CaCl_2$, 20mM $MgSO_4$, 10% glycerol. Aliquots of cells were snap frozen in liquid nitrogen before storage at -70°C.

2.4.2.4. Transformation of thermocompetant bacteria

10µl of ligation reaction mixture or 100ng of whole plasmid DNA was mixed on ice with 40µl of thermocompetant bacteria suspension (see above) and incubated on ice for 30 mins. The DNA/cell mixture was heat-shocked at 42°C for 2 mins before returning to ice for a further 15 mins. The cell suspension is resuspended in L-broth/20mM glucose, incubated at 37°C and spread on L-agar plates as described above for electrocompetant cell tranformation.

2.4.2.5. Generated plasmid constructs

2.4.2.5.1. SUMO-1 (1-97) mutants

SUMO-1 mutations were produced using a single stage PCR method. 20ng of template WT-SUMO-$1_{(1-97)}$-pGEX-2T (13), 100ng of each primer and New England Biolabs Vent DNA Polymerase were used for PCR reactions. For cloning into pcDNA-3 and pGEX-2T vectors the 5′ primer (5′-ACAAACGGATCCATGTCTGACCAGGAGGCCAAA-3′) containing a BamHI restriction site was used, while for cloning into the pET-3a vector the 5′ primer (5′-ACAAACCATATGTCTGACCAG-GAGGCAAAA-3′) containing an NdeI restriction site was used. For each mutation a different 3′ primer containing a BamHI restriction site was used: E93R-Q94L-T95R-SUMO-$1_{(1-97)}$ (RLR-SUMO-$1_{(1-97)}$); 5′-ATAATAGGATCCTCAACCCCCCCTTAGTCGCTGATAAAC-3′, E93R-

T95R-SUMO-$1_{(1\text{-}97)}$ (RR-SUMO-$1_{(1\text{-}97)}$); 5′-ATAATAGGATCCTCAACCCC-CCCTTTGTCGCTGATAAAC-3′, T95R-SUMO-$1_{(1\text{-}97)}$; 5′-ATAATAGGATCC-TCAACCCCCCCTTTGTTCCTG-3′, Q94L-SUMO-$1_{(1\text{-}97)}$; 5′-ATAATAGGATC-CTCAACCCCCCGTTAGTTCCTGATAAAC-3′, E93R-SUMO-$1_{(1\text{-}97)}$; 5′-AT-AATAGGATCCTCAACCCCCCGTTTGTCGCTGATAAAC-3′.

All SUMO-1 mutant DNAs were PCR amplified using the BamHI forward primer and the specific reverse mutant primers, while T95R-SUMO-1 was additionally amplified using the NdeI forward primer and the T95R mutant reverse primer. PCR products were gel extracted before digestion with either BamHI alone (for cloning into pcDNA-3 or pGEX-2T) or with NdeI and BamHI (for the cloning of T95R-SUMO-1 into pET-3a). All SUMO-$1_{(1\text{-}97)}$ mutant DNAs were cloned into the N-terminal HA-Tag pcDNA-3 vector for transient cell transfection. Similarly all SUMO-1 mutant DNAs were cloned into the pGEX-2T expression vector, except for T95R-SUMO-1 which was cloned into the pET-3a vector. This was necessary as the T95R-SUMO-1 mutant could not be expressed using the pGEX-2T plasmid in *E. coli* B834.

2.4.2.5.2 SUMO-1 (1-101) mutants

All SUMO-1 mutants coding for 101 amino-acid protein products were PCR amplified as described above for the 1-97 mutants, with the substitution of 3′ primers containing 4 codons of the additional C-terminal residues and a BamHI restriction site: E93R-Q94L-T95R-SUMO-$1_{(1\text{-}101)}$ (RLR-SUMO-$1_{(1\text{-}101)}$); 5′-ATAATAGGATCCTCAAACTGTTGAATGACCCCCCC-

TTA-3', E93R-T95R$_{(1-101)}$-SUMO-1 (RR-SUMO-1$_{(1-101)}$); 5'-ATAATAGGATC-CTCAAACTGTTGAATGACCCCCCCTTTGTCGCTGATAAA-3', T95R-SUMO-1$_{(1-101)}$; 5'-ATAATAGGATCCTCAAACTGTTGAATGACCCCCCC-TTTGTTCCTG-3', Q94L-SUMO-1$_{(1-101)}$; 5'-ATAATAGGATCCTCAAACTG-TTGAATGACCCCCCCTTAGTTCCTGATAAA-3', E93R-SUMO-1$_{(1-101)}$; 5'-ATAATAGGATCCTCAAACTGTTGAATGACCCCCCCGTTTGTCGCTGA-TAAA-3'. All SUMO-1$_{(1-101)}$ mutant DNA PCR products were cloned into the N-terminal HA-Tag pcDNA-3.

2.4.2.5.3. SUMO-2 and SUMO-3

The full-length (FL) cDNA clones of SUMO-2 (309 nucleotides) and SUMO-3 (285 nucleotides) were cloned by Reverse-transcriptase polymerase-chain reaction (RT-PCR) using the "Titan™" one tube RT-PCR system (Boeheringer Mannheim) as per the manufacturers instructions, using HeLa Poly A+ RNA as template (a gift from David Hay University of St. Andrews). SUMO-2 was cloned using primers (5'-TCCCCGCGCCGCTC-GGAATTCATGTCCGAG-3' and 5'-CCCGAATTCGGGACGGGCCCTCTAG-AAACT-3'), and SUMO-3 was cloned using primers (5'-GAGGAGACTCC-GGCGGGATCCATGGCCGACGAA-3' and 5'-GTAGAATTCCAGGTTCCC-TTTTCAGTAGAC-3'). Shorter DNA constructs which code for proteins terminated at the diglgycine (276 nucleotides for SUMO-2-GG, and 279 nucleotides for SUMO-3-GG) were amplified by PCR using the single downstream primer (5'-CCCGAATTCCTAACCTCCCTGCTGCTGTTGGA-ACAC-3') for both, with the respective upstream primer as described

above. The restriction sites within the primers allowed the cleavage of the cDNAs by the enzymes BamHI and EcoRI, and subsequent ligation into both pcDNA-3 and pGEX-2T.

2.4.2.5.4. Ubc9 mutants

Mutants of Ubc9 encoding combinations of changes of amino-acids Asp 100 and Lys 101 to Ala were generated using two-stage PCR reactions. All mutants were PCR amplified using the external primers; 5'-ACAAAC-GGATCCATGTCGGGGATCGCCCTCAGC-3' and 5'-GCCGCGGAATTCTT-ATGAGGGCGCAAACTTCTTGGC-3'. Each mutant was PCR amplified using the following internal primers: D100A-K101A-Ubc9 (DK-AA-Ubc9); 5'-CCTCCAGTCCGCGGCCTCCTCTAA-3' and 5'-TTAGAGGAGGCCGCGG-ACTGGAGG-3', D100A-Ubc9; 5'-CCAGTCCTTGGCCTCCTCTAA-3' and 5'-TTAGAGGAGGCCAAGGACTGG-3', K101A-Ubc9; 5'-CCTCCAGTCCGCG-TCCTCCTC-3' and 5'-GAGGAGGACGCGGACTGGAGG-3'. The external primers facilitate cloning of PCR products into pGEX-2T using the restriction enzymes BamHI and EcoRI.

2.4.3. DNA sequencing

All DNA constructs were verified by automated DNA sequencing on an ABI PRISM^{TM} 377 DNA Sequencer (St. Andrews University DNA sequencing unit).

2.5. Expression and Purification of Unlabelled Recombinant Proteins

GST-WT-Ubc9, GST-Ubc9 mutants, GST-SUMO-2-FL, GST-SUMO-2-GG, GST-SUMO-3-FL, GST-SUMO-3-GG, GST-WT-SUMO-$1_{(1-97)}$ and GST-SUMO-$1_{(1-97)}$ mutants were expressed and purified from *E. coli* strain B834 as described previously (Jaffray *et al.*, 1995). Recombinant GST-SAE2/SAE1 was a gift from Joana Desterro (University of St. Andrews). All GST-SUMO proteins were cleaved by thrombin while bound to glutathione-sepharose beads. For freeze-drying, proteins were dialysed against 20mM ammonium bicarbonate, 1mM dithiothrietol (DTT), before concentration calculations. Lyophilised protein samples were taken up in 50mM Tris/HCl pH 7.5, 1mM DTT to a concentration of $10mg.ml^{-1}$.

T95R-SUMO-1-pET-3a was used to transform BL21 de3 *E. coli* and expressed and extracted as described previously (Jaffray *et al.*, 1995). The protein was purified by HPLC anion-exchange chromatography and Superdex 75 size exclusion chromatography. Protein concentration, lyophilisation and resolublization were carried out as described above. Both full length and GG SUMO-2 and SUMO-3 WT proteins, and all SUMO-1 and Ubc9 mutant protein masses were verified by MALDI TOFF mass spectrometry before use in the assays.

2.6. Expression and Purification of $^{15}N/^{13}C$ Double Labelled SUMO-1 for NMR Analysis

BL21 *E. coli* bacteria were transformed to ampicillin resistance with pGEX-2T-WT-SUMO-$1_{(1-97)}$. Transformed BL21 cells were streaked on minimal medium-agar plates; 47.8mM Na_2HPO_4, 22mM KH_2PO_4, 8.5mM NaCl, 18.8mM NH_4Cl, 1mM $MgSO_4$, 22.2mM glucose, with 100µg/ml^{-1} ampicillin and 2% agar. A single, transformed bacterial colony best adapted to minimal medium growth conditions was used to innoculate 10ml minimal medium (as above without agar) exchanging the NH_4Cl and glucose with 98% ^{15}N NH_4Cl and 99% ^{13}C glucose (Fluka). This culture was incubated at 37°C with agitation until the optical density at 600nm was approximately 0.5. This culture was used to innoculate 500ml of the same medium which was incubated at 37°C with agitation until the optical density at 600nm was 0.6. The culture was IPTG induced and the protein purified as detailed above for SUMO-1.

2.7. Quantitation of Protein Concentration

For purified SUMO-1, SUMO-2 and SUMO-3 WT constructs protein concentrations were determined using calculated ε_{280} extinction coefficients and the absorbance at 280nm and Beer's Law; $A=\varepsilon Cl$. For all other proteins, concentration was assayed using the Bradford's method (Bradford, 1976). Protein samples were mixed with Bradford's reagent (Biorad) and the absorbance at 595nm measured. Absorbances were used to

generate a standard curve by using varying concentrations of bovine serum albumin (BSA).

2.8. SDS PAGE and Western Blot Analysis

For analysis of all protein samples other than COS7 cell extracts, protein solutions were resuspended in sample buffer (1X: 20mM Tris/HCl pH6.8, 2% SDS, 5% β-mercaptoethanol, 2.5% glycerol and 2.5% bromophenol blue) and denatured at 100°C for between 2 and 5 mins. Proteins were fractionated by electrophoresis on 8%, 10%, 12% and 15% polyacrylamide gels containing SDS (as required), using mini-gel equipment as described by the manufacturer (Bio-Rad). NEB protein molecular weight markers (prestained and non-prestained) were used to gauge apparent molecular weights where necessary. Gels were then either Coomassie blue stained (0.2% Coomassie brilliant blue R250, 50% methanol, 10% glacial acetic acid) for 20-30 min and destained (20% methanol, 10% glacial acetic acid), or transferred to polyvinylidene difluoride (PVDF) membrane using a wet blotter (Bio-Rad). Membranes were blocked with phosphate buffered saline (PBS), 5% skimmed milk powder, and 0.1% Tween 20 for at least 1h, before incubating with required "primary" antibody diluted in the same buffer for 1h. Horseradish peroxidase conjugated antibodies were used as "secondary" (see above for all antibody types). Secondary antibody detection was performed using an ECL system.

For analysis of transiently transfected COS7 cells, extracts were lysed in SDS sample buffer (5% SDS, 0.15M Tris-HCl pH 6.7, 30% Glycerol) diluted 1:3 in RIPA buffer (25mM Tris pH 8.2, 50mM NaCl, 0.5% Nonidet P-40, 0.1% SDS), containing 10mM iodoacetamide and complete protease inhibitor cocktail (Boehringer Mannheim). Lysates were sonicated briefly, and cleared by centrifugation. Proteins were fractionated by electrophoresis and analysed as described above.

2.9. NMR Analysis of SUMO-1

Lyophilised SUMO-1 was dissolved in 100mM Phosphate buffer pH7.5 1mM sodium dithionite 10% D_2O, to a concentration of approximately 0.8 mM. Dissolved material was analysed as described by (Grzesiek *et al.*, 1992; Sklenar *et al.*, 1993) in a Varian Unity-Plus 500 spectrometer with a field strength of 500.13 MHz at a temperature of 30.0°C. HSQC spectra contained 512 complex points in f2, and either 128 or 64 complex points in f1. HNCA and HN(CO)CA spectra contained 512 complex points in f3, 16 in f1 and 48 in f2. All HSQC, HNCA and HN(CO)CA spectra were zero filled once in each dimension, and apodised by a sine-bell square window function shifted by $\pi/2$, prior to Fourier transform. Strong residual water signals were removed using WATERGATE solvent suppression (Piotto *et al.*, 1992). All data were processed and presented using Varian VNMR software on a Silicon Graphics $SGIO_2$.

2.10. Resolublisation and Refolding of Precipitated SUMO-1

Precipitated SUMO-1 was recovered to NMR structural analysis grade by employment of resolublisation and refolding techniques. Precipitated double labelled SUMO-1 was pelleted out of suspension by centrifugation, and the pellet resolublised and denatured to a concentration of 10mg.ml^{-1} in 6M guanidine hydrochloride. The protein in this solution was refolded by injection into a solution of 1M NDSB-201 (Fluka) as described previously (Goldberg *et al.*, 1995), 5mM dithiothrietol, 1mM sodium EDTA at a ratio of 1:9 while vortexing at 4 °C. The resultant solution of SUMO-1 (1mg.ml^{-1}) was dialysed against 20mM ammonium bicarbonate 1mM dithiotrietol and lyophilized as described above.

2.11. ^{35}S-Methionine Labelling of Proteins

^{35}S-labelled proteins were generated using *in vitro* transcription/translation from 1µg of plasmid DNA and a wheat germ coupled transcription/translation system with ^{35}S-methionine (Amersham), according to the instructions provided by the manufacturer (Promega).

2.12. ^{125}I Labelling of Proteins

The recombinant mutant and WT SUMO-1 proteins were radiolabelled with carrier-free Na ^{125}I (Amersham) by the chloramine-T method (Ciechanover et al., 1980). Free ^{125}I was removed from solution by dialysis against 50mM Tris pH7.5, 1mM DTT. The remaining labelled proteins were fully reduced by the addition of 100mM DTT followed by dialysis to 10mM iodoacetamide and 10mM DTT, then to 10mM iodoacetamide only, followed finally to 20mM ammonium bicarbonate, 1mM DTT in preparation for freeze-drying and resolublising as described above.

2.13. Preparation of HeLa Cell Fractions

Q-Sepharose fractions containing the SUMO-1-activating enzyme (SAE) activity were prepared from HeLa cell extracts as described (Desterro *et al.*, 1998). Homogeneous SAE was prepared from a pool of Fr II.3 and II.4 by affinity chromatography on immobilised SUMO-1 as described previously (Desterro *et al.*, 1999). The enzymatic activity present in the purified samples was determined by thioester assay as described below, and protein purity was assessed by SDS PAGE followed by Coomassie staining.

2.14. *In vitro* SUMO Thioester Assay (^{125}I-labelled SUMO-1)

Samples of the ^{125}I-labelled SUMO-1 proteins were subjected to SDS-PAGE (15%) and dried gels analysed by phosphorimaging. This data was

used to standardise the activities of the proteins for thioester analysis. Equal amounts of ^{125}I-labelled proteins were assayed for their ability to form thioester bonds with approximately 100ng of purified SAE and 1µg Ubc9 in a 10µl reaction volume as described (Desterro *et al.*, 1997). Reactions were incubated for 30 min at 30°C and terminated by incubating for a further 30 min at 30°C with 2% (w/v) SDS, 4M Urea. Samples were fractionated by SDS PAGE (15%) such that the separation between the SUMO-1~SUMO-1 disulfide dimers and the SUMO-1~Ubc9 thioester was maximised. Dried gels were analysed by phosphorimaging (Fujix BAS100, MacBas Software).

2.15. *In vitro* SUMO Thioester Assay (Unlabelled SUMO-1/-2/-3)

10µg of each SUMO protein construct was incubated with 1.2 µg recombinant Ubc9 in a 10µl reaction volume of 50mM Tris/HCl pH 7.5, 10mM $MgCl_2$, 1mM ATP, 100U.ml^{-1} Inorganic pyrophosphatase (IPP), in the presence or absence of 1.4µg of recombinant SAE1/SAE2. Reactions were incubated for 30 min at 37°C and terminated by incubating for a further 30 min at 30°C with 2% (w/v) SDS, 4M Urea. Samples were fractionated by SDS PAGE (15%) to maximise the separation between the Ubc9~Ubc9 disulfide dimers and the SUMO-1~Ubc9 thioester. Gels were subsequently Western-blotted and analysed with antibodies specific to Ubc9, SUMO-1 or SUMO-2/-3 as described above.

2.16. *In vitro* SUMO-1/-2/-3 Conjugation Assay

To test the ability of particular SUMO protein constructs to conjugate to ^{35}S-methionine labelled *in vitro* transcribed/translated p53, IκBα, HDac4, Adenovirus pV and PML, an *in vitro* conjugation assay was used. 1µl of ^{35}S-substrate was incubated with either 2µl of HeLa cell fraction containing SUMO-1 E1 activity (Fr.II.3), or 350ng recombinant SAE1/2, in a 10µl reaction including an ATP regenerating system (50mM Tris pH 7.5, 5mM $MgCl_2$, 2mM ATP, 10mM creatine phosphate, 3.5U.ml^{-1} creatine kinase and 0.6U.ml^{-1} IPP, 1mg.ml^{-1} SUMO-1, 600ng Ubc9. Reactions were incubated at 37°C for 2h. For assay of Ubc9 mutants WT-Ubc9 was substituted with the appropriate proteins in the reaction outlined above. Reactions were terminated by the addition of SDS sample buffer containing β-mercaptoethanol, and reaction products were fractionated by SDS PAGE (8-10%) and dried gels analysed by phosphorimaging as described above.

2.17. *In vitro* SUMO-1 Maturation Assay

To analyse the ability of SUMO-1 mutants to be cleaved by C-terminal hydrolase activities, full-length ^{35}S-SUMO-1 proteins were *in vitro* translated from pCDNA-3-HA constructs, and incubated for varying lengths of time at 30°C, in reactions containing 1:2:7 ratio volumes of ^{35}S-SUMO-1 in vitro transcription/translation protein solution, crude nuclear HeLa cell extract (from Prof. R. T. Hay, university of St. Andrews) known

to contain both SUMO-1-specific C-terminal hydrolase and isopeptidase activities, and 50mM Tris pH 7.5, 1mM DTT. Assays were terminated by addition of SDS-β-mercaptoethanol sample buffer, and fractionated by SDS poly-acrylamide gel electrophoresis, before drying and analysis by phosphorimaging.

2.18. *In vitro* SUMO-1 Deconjugation Assay

To test the susceptibility of SUMO-1 variants for removal from protein ^{35}S-protein substrates by isopeptidase activities, pools of conjugated proteins were firstly generated using the conjugation assay described above, incubated for 3h. Conjugation reactions were then inhibited by the addition of 10mM EDTA, and SUMO-1 protease-containing crude HeLa nuclear extract (see above) added to a ratio of 1µl per 5µl reaction. Deconjugation reactions were incubated at 30°C for 60 mins. Assays were terminated by addition of SDS-β-mercaptoethanol sample buffer, and fractionated by SDS poly-acrylamide gel electrophoresis, before drying and analysis by phosphorimaging.

2.19. Ubc9 Binding Assays

GST, GST-WT-Ubc5, GST-WT-Ubc9 or GST-Ubc9 mutants were bound to glutathione sepharose beads and analysed by SDS polyacrylamide gel electrophoresis to standardise the mass:volume ratio for each. Equal masses of protein (between 5 and 20 µl bead volume) were incubated for 2h

at room temperature in 100µl binding buffer (20mM Tris/HCl pH7.5, 100mM NaCl, 2mM EDTA, 0.1% NP-40, 2mM DTT, 0.05% BSA, 5% glycerol) containing 10µl ^{35}S-PML or ^{35}S-p53 produced by *in vitro* transcription translation. Beads were washed once with 150mM NaCl, 5mM EDTA, 50mM Tris/HCl pH 7.5, 0.5% NP-40, then once with the same solution omitting the NP-40. Beads were centrifuged and resuspended in 20µl 1X β-mercaptoethanol/SDS sample buffer, before boiling and fractionation by SDS PAGE. Gels were dried and analysed by phosphorimaging.

2.20. Cell Culture and Transfections

Human ovarian cancer (HeLa) cells and (African Green Monkey kidney) COS7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS).

For Western-blot analysis of SUMO-1/-2/-3 proteins, sub-confluent COS7 cells were transfected with 1µg of plasmid DNAs for 14h in 6-well plates using Lipofectamine (Gibco). Medium was replaced with fresh DMEM + 10% FCS, and incubated for a further 24 to 36h.

For fluorescence microscopy of SUMO-1 mutants, HeLa cells were transfected by electroporation (950mF, 200V, Equibio Easyject plus) as described (Arenzana-Seisdedos *et al.*, 1997). A total of 10µg of plasmid DNA encoding the HA-SUMO-1-WT/mutants were transfected into 5 x 10^6 cells. After transfection, cells were seeded into 12 well plates containing glass

microscope coverslips, and incubated for 24h. Expression of HA-SUMO-1 and the various mutants were verified by anti-HA Western-blot as described below.

2.21. Indirect Immunofluorescence Analysis of HeLa Cells

Transfected HeLa cells adhered to glass microscope coverslips were fixed with 3% paraformaldehyde in PBS for 10 min. Following 3 washes with PBS and two 10 min incubations with 100mM glycine in PBS, cells were permeablised with a 0.2% solution of Triton-X100 in PBS for 10 min. After a further 3 PBS washes, cells were blocked using a PBS solution containing 5% donkey serum (Diagnostics Scotland). Expressed HA-tagged SUMO-1-WT/mutants were detected with the anti-HA polyclonal antibody, and endogenous PML nuclear bodies were detected using the mouse mono-clonal (both described above). Primary antibodies were diluted in blocking buffer and incubated on permeablised cells for 1h at room temperature. Following three washes with blocking solution, secondary antibodies, goat anti-mouse FITC and goat anti-rabbit Texas-red conjugates, were incubated with cells for 45 min at room temperature. Cells were washed 3 times with PBS before coverslips were affixed to glass microscope slides using Mowiol (Calbiochem). Fluorescence microscopy was carried out with a Nikon Microphot-FXA microscope. Images were captured with a SPOT CCD camera and manipulated using Photoshop 6.0 software.

3. RESULTS & DISCUSSIONS

3.1. 3-Dimensional Structure of SUMO-1

3.1.1. Results

Biochemical studies in 1997 of the Ub-conjugating enzyme Ubc9 revealed that it was not Ub-specific, but was in fact the E2 enzyme for the Ub-like protein SUMO-1 (Desterro *et al.*, 1997). Furthermore, structural studies revealed distinct electrostatic and minor topological differences between Ubc9 and Ub E2s which could explain their affinity for different protein modifiers (Giraud *et al.*, 1998; Tong *et al.*, 1997). Although SUMO-1 and Ub share only 18% sequence identity (shown in Figure 3 and Table 3) predicted 3D structures of SUMO-1 based on Ub revealed that they may share the same protein fold (Giraud M-F., personal comm.). To aid investigations into the associations with different conjugating enzymes and substrates we sought to resolve the solution-state 3D structure of SUMO-1 using NMR spectroscopy.

3.1.1.1. Optimisation of SUMO-1 protein sample and NMR analysis conditions

Primary one-dimensional proton spectra using recombinant purified SUMO-$1_{(1\text{-}97)}$ (exposed C-terminal glycine form) from *E. coli* , with atmospheric abundances of magnetic nuclei, showed good chemical shift separation. However, overlapping peaks in 2D and 3D experiments meant that the structure could not be resolved using this protein sample (data not

shown). Consequently, a sample of SUMO-1 97-99% isotopically enriched in ^{15}N and ^{13}C was produced using BL21 *E. coli* bacteria grown in M9 minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C D-glucose, as described in the Materials and Methods section. Initially, a lyophilized sample of $^{15}N/^{13}C$ SUMO-1 was dissolved in 100mM phosphate buffer pH 5.5, 1mM sodium dithionite 10% D_2O to a protein concentration of ~0.8mM, and analysed in a Varian Unity-Plus 500 spectrometer at a temperature of 45°C. At this pH and temperature although proton exchange was reduced and protein tumbling time increased, double labelled SUMO-1 aggregated very rapidly, and almost half the sample precipitated within 12h. Optimization trials using smaller aliquots of double labelled SUMO-1 revealed that the protein was much more stable at pH7.5 and 30°C, taking more than 48h for visible signs of precipitation to appear. Under these conditions preliminary data was obtained using fresh samples of $^{15}N/^{13}C$ SUMO-1.

3.1.1.2. Precipitated SUMO-1 can be recovered to native conformation using NDSB-201 refolding procedures

Unfortunately, given the higher acquisition time for 3D spectra, line-broadening due to protein aggregation affected NOESY-HSQC, COSY-HN(CO)CA and COSY-HNCA experiments, resulting in low quality 2D and 3D spectra. Subsequently either fresh protein was to be prepared, or the original precipitated material recovered. Using 6M guanidine hydrochloride resolublisation and a non-detergent sulfobetane NDSB-201

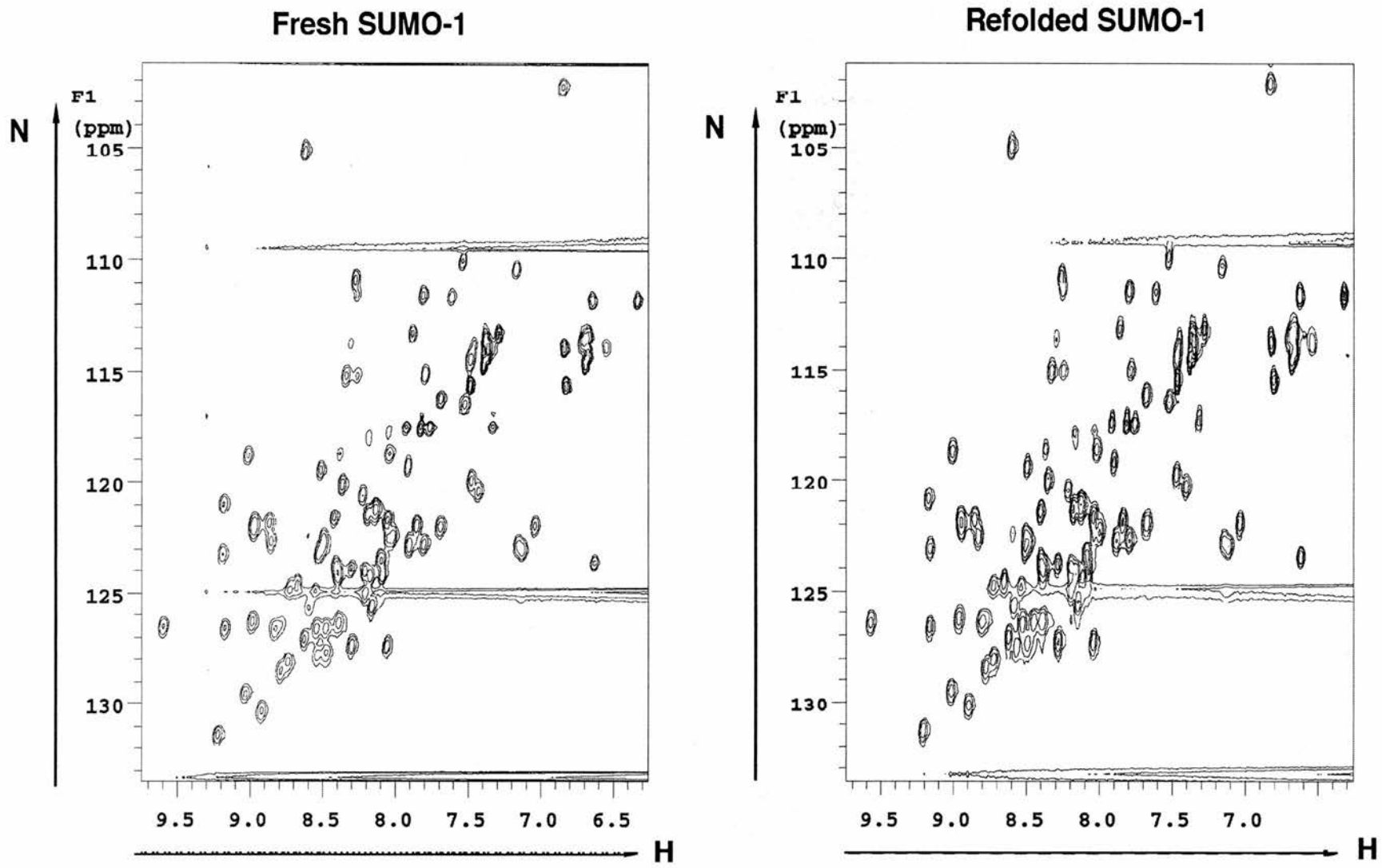


Figure 5. 2DNOESY-HSQC NMR spectra of both freshly prepared, and refolded samples of $^{15}N/^{13}C$ labelled SUMO-1. Data obtained and processed as described in the Materials and Methods. Nitrogen and hydrogen dimensions are shown.

refolding method (Goldberg *et al.*, 1995), precipitated $^{15}N/^{13}C$ SUMO-1 could be recovered (as described in Materials and Methods). Figure 5 shows that NOESY-HSQC spectra for both fresh and refolded SUMO-1 are essentially identical, indicating that the resolublisation and refolding method efficiently recovers precipitated protein to native conformation. Consequently, samples of $^{15}N/^{13}C$ SUMO-1 were analysed and recycled until adequate data was obtained.

3.1.1.3. COSY-HNCA and HN(CO)CA spectra of SUMO-1 are of high enough resolution to make primary α-carbon assignments

Figure 6 shows both HN(CO)CA and HNCA spectra for SUMO-1 with the proton dimension compressed into the page. The spectra show good separation between peaks, and peak-picking algorithms define 92 individual shifts from the 97 (+ 2) α-carbons of the protein (data not shown) (two extra residues, glycine and serine are added to the N-terminus of pGEX-2T expressed protein after cleavage of the GST tag by thrombin). "1" and "1'" indicated in Figure 6, are shifts from the α-carbons of adjacent amino-acids, and were used as a starting point for the backbone resonance assignments.

Resonances were assigned for approximately 60 of the 99 α-carbons before the publication in February 1998 of the NMR resolved 3D structure of SUMO-1 (Bayer *et al.*, 1998). Bayer *et al* also had problems with aggregation and line-broadening, and eventually performed NMR experiments at a similar temperature and pH to

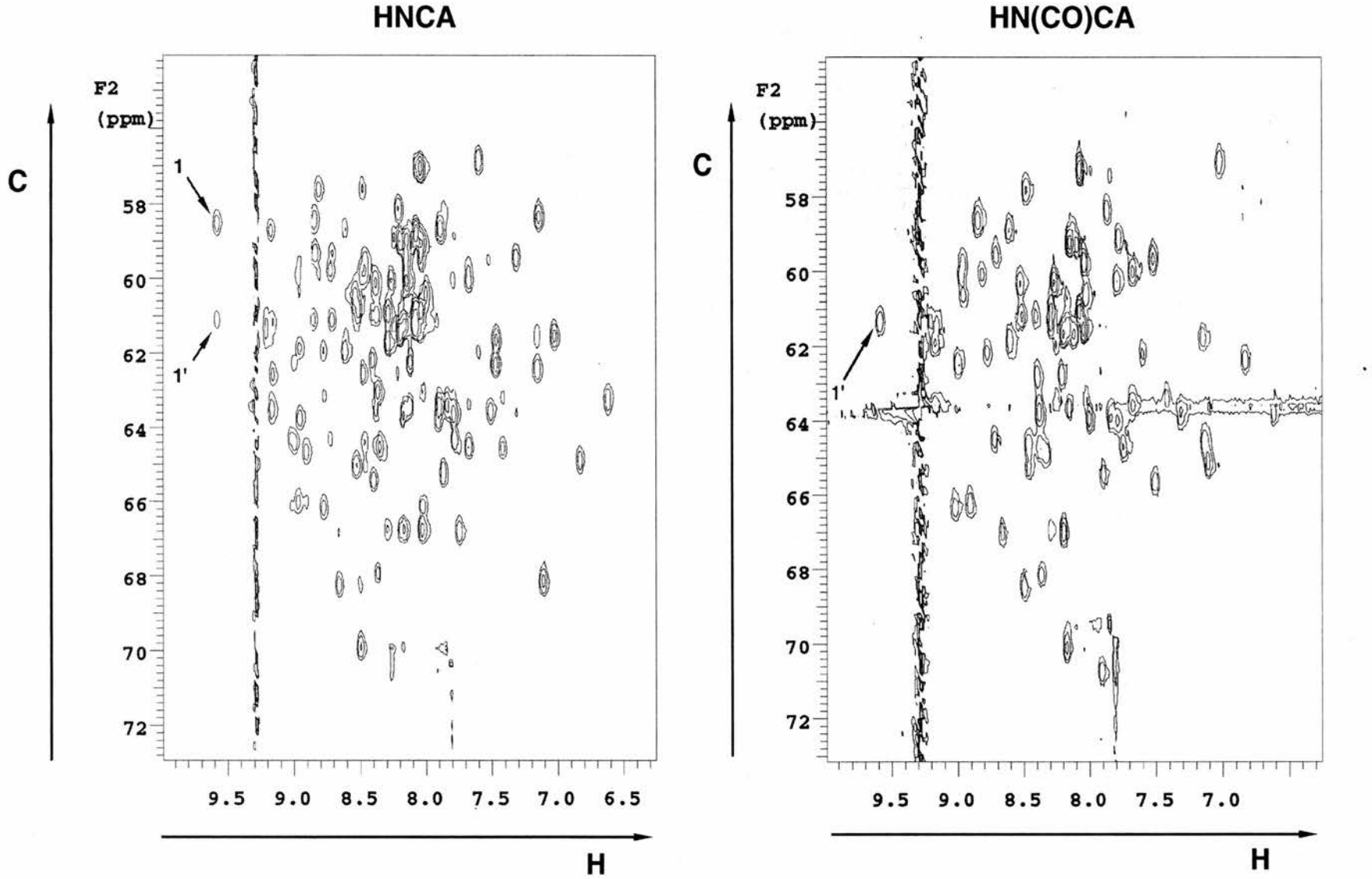


Figure 6. 3DCOSY-HNCA and HN(CO)CA NMR spectra of $^{15}N/^{13}C$ labelled SUMO-1. Data obtained and processed as described in the Materials and Methods. Carbon and hydrogen dimensions are indicated, and the nitrogen dimension is compressed into the plane of the page. 1 and 1' indicate resonances of α-carbons from adjacent amino-acids.

ourselves (27°C and pH7.0). In contrast to data presented here with $^{15}N/^{13}C$ double-labelled SUMO-$1_{(1\text{-}97)}$, Bayer *et al* used the full-length construct of SUMO-1 (1-101) with the His/Ser/Thr/Val C-terminal peptide attached, and used only ^{15}N single-labelled SUMO-1. Bayer *et al* also took a different experimental approach to making resonance assignments, and resolving the secondary and tertiary structures of SUMO-1. They used 2D DQF-COSY and 2D clean-TOCSY to obtain spin systems, and 3D ^{15}N-TOCSY-HMQC spectra to clarify ambiguous overlaps. They obtained sequential connectives from 2D NOESY spectra, allowing them to make resonance assignments. Short and medium range NOEs were used to define secondary structure, which was generally confirmed using the chemical shift index (CSI) method (Wishart *et al.*, 1992). A family of 30 3D structures were calculated by simulated annealing (SA), followed by refinement using X-PLOR 3.1 and 3.854 (Brunger, 1993; Brunger, 1996). The remainder of this section discusses the resulting Bayer SUMO-1 structure.

3.1.2 Discussion

3.1.2.1. SUMO-1 features a ubiquitin superfold

Primary and secondary structure alignments of Ub and SUMO-1 (Figure 7A) show that although only 18% identical in amino-acid sequence, the two proteins have almost identical secondary structures. The SUMO-1 tertiary structure consists of a five-stranded β-sheet with two α-helices, with helix α1 (Leu44 to Gln55) rotated approximately 45° relative to the first β-sheet, which represents a typical Ub superfold (Figure 7B). The Ub superfold is a good example of how proteins with insignificant sequence identity can form proteins of the same 3D structure, and has been classed as one of nine superfolds identified by structural homology (Orengo *et al.*, 1994). Numerous proteins from diverse functional backgrounds and species are known to contain a Ub-superfold. Examples include the ferredoxin protein from blue-green algae, the B2 immunoglobulin-binding domain of streptococcal protein G, and the Ras-binding domains of the eukaryotic Raf kinase, RBD (Nassar *et al.*, 1995), RalGDS (Geyer *et al.*, 1997), and myr5 RBD (Kalhammer *et al.*, 1997). 3D protein back-bone overlays of Ub and SUMO-1 (Figure 8.A) show that the two structures closely resemble each other. In fact, the rms separation value for the two core structures (residues 21-93 in SUMO-1 and 1-72 in Ub) is only 2.1Å (Bayer *et al.*, 1998). Furthermore, the C-terminal glycine (97) involved in the formation of thioester bonds with the SUMO-

conjugation machinery, and isopeptide bond with the substrate is in the same position as the corresponding residue in Ub (Gly 76), relative to the Ub fold.

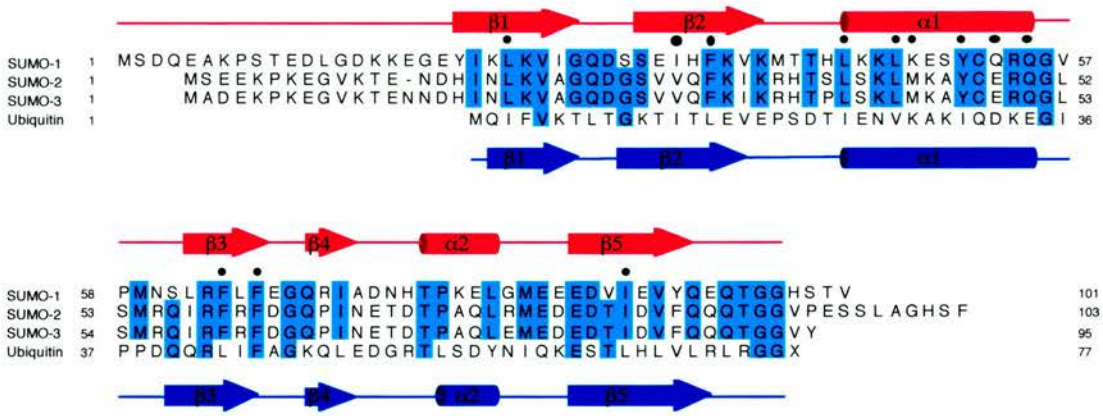
Both α1 helices in SUMO-1 and Ub are strongly amphipathic, with solvent exposed hydrophilic residues on one side, and buried hydrophobic residues on the other. Contacts at the helix-sheet interface in SUMO-1 occur between residues Leu24, Ile34, Phe36, Phe64, Phe66 and Ile88 on the β-strands, and Leu44, Leu47, Lys48, Tyr51, Gln53, and Gln55 from the α1 helix. Of these 12 hydrophobic core residues 3 are identical and 7 homologous to residues in Ub, and are thus thought to be essential to maintain the Ub fold. Interestingly, although SUMO-1 only shares 52% sequence identity with SUMO-2 and SUMO-3, 9 of these residues are identical, and the remaining 3 homologous to residues in SUMO-2 and SUMO-3 (Figure 3), lending considerable weight to the hypothesis that the other two members of the SUMO family are likely to be of very similar structure to SUMO-1.

3.1.2.2. Significant structure and charge differences between SUMO-1 and ubiquitin may reflect their differing protein associations, or functional consequences to substrates

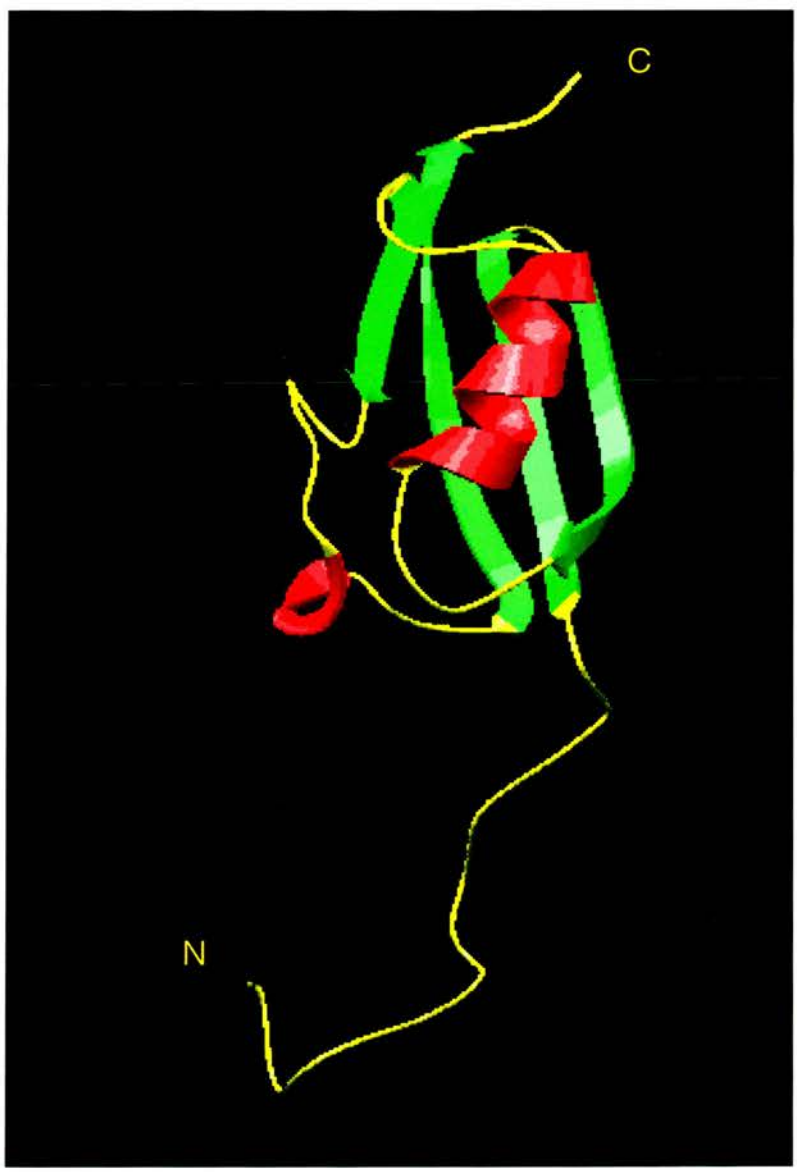
Although Ub and SUMO-1 are almost identical in core protein structure, several significant electrostatic and structural differences can been seen between the two proteins. The major structural difference between them is a flexible 20 residue N-terminal extension in SUMO-1

Figure 7. Primary and secondary structure alignments of Ub and SUMO-1, and ribbon representation of the 3D structure of SUMO-1. **A.** Primary sequence alignment of Ub, SUMO-1, SUMO-2 and SUMO-3, with secondary sequences of Ub (below in blue) and SUMO-1 (above in red). Identical residues are shown in cyan, and residues in α-helix 1 and the β-strands thought to be essential for the integrity of the Ub-fold structure, are marked with an black dot. **B.** A ribbon representation of the 3D structure of SUMO-1, the C-terminus and the extended N-terminus are indicated. Images Produced using the Swiss PDB viewer package v3.6b (Guex *et al.*, 1999) (A. and B. adapted from (Bayer *et al.*, 1998)).

A


β1 β2 α1
SUMO-1 1 MSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGV 57
SUMO-2 1 MSEEKPKEGVKTE-NDHINLKVAGQDGSVVQFKIKRHTSLSKLMKAYCERQGL 52
SUMO-3 1 MADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGL 53
Ubiquitin 1 MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGI 36
β1 β2 α1
β3 β4 α2 β5
SUMO-1 58 PMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGGHSTV 101
SUMO-2 53 SMRQIRFRFDGQPINETDTPAQLRMEDEDTIDVFQQQTGGVPESSLAGHSF 103
SUMO-3 54 SMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGVY 95
Ubiquitin 37 PPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGX 77
β3 β4 α2 β5

B


C
N

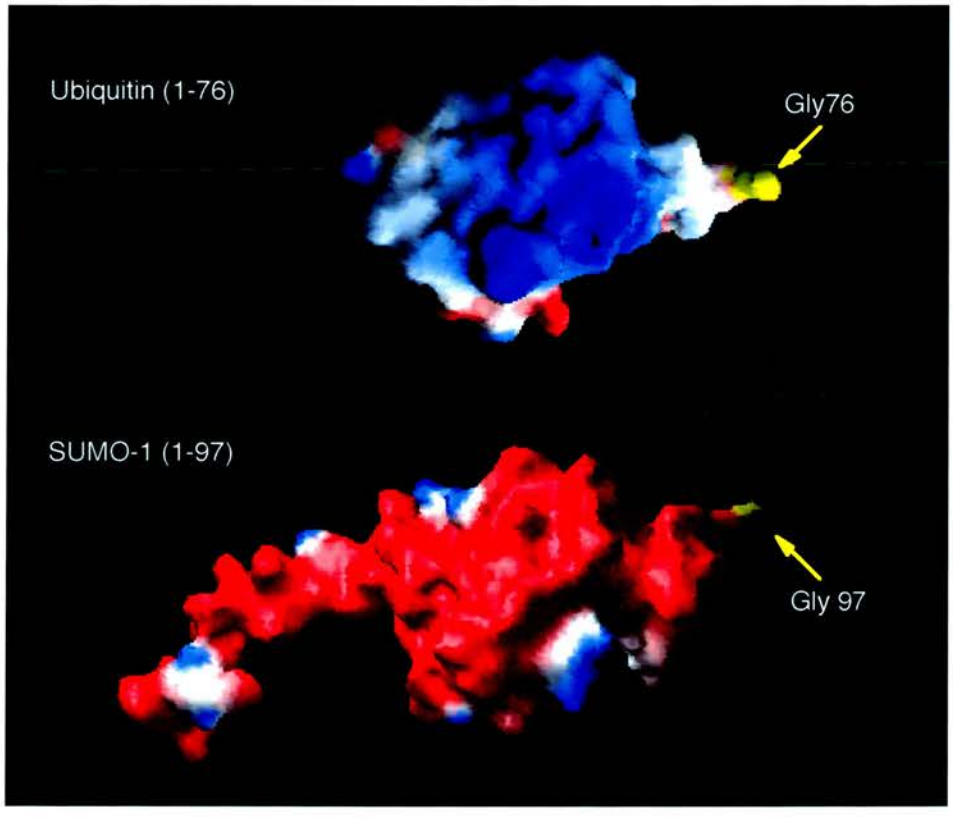
(Bayer *et al.*, 1998), which is apparently conserved as 16 residue extensions in the other two members of the SUMO family (Figure 3). The second major difference between the SUMO-1 and Ub is their contrasting surface charge distributions. The isoelectric points of SUMO-1 (~pH 5.2) is significantly more acidic than Ub's, which is almost neutral (~pH 6.9). These values are reflected in the differing electrostatic surface potentials of the two proteins, which in the case of Ub is mostly composed of neutral or basic regions, while the surface of SUMO-1 is predominantly acidic (Figure 8.B). One such negatively charged region in SUMO-1 is formed by Glu83, Glu84, Glu85 and Asp86, of which only Glu83 is conserved in Ub (Glu64). Together with negatively charged residues from N-terminal tail (Glu11, Asp12, Glu15, Glu18 and Glu20), these residues form a large contiguous patch of negativity on the surface of SUMO-1, which is not present in Ub. Subsequently, given that SUMO-1 and Ub are metabolised by different enzymes (Schwienhorst *et al.*, 2000), it is tempting to suggest that this area of negative charge may be involved in SUMO-1-specific protein-protein interactions. Interestingly, conjugation efficiency studies of a SUMO-1 protein with an N-terminal deletion (ΔN21-SUMO-1 protein) and SUMO-1 mutations of Glu83, Glu84 and Glu85, have shown little or no conjugation defects *in vitro* or *in vivo* (Hodges. M. personal comm.). However, immunofluorescence studies by Hodges, M. *et al* with the ΔN21-SUMO-1 protein, showed that it did not present the characteristic diffuse nuclear staining of WT-SUMO-1, but was only visible in PML NBs, the predominant WT-SUMO-1 containing nuclear structures (Boddy *et al.*, 1996; Duprez *et al.*, 1999; Ishov *et al.*, 1999).

Figure 8. Superimposition of the backbone structures of Ub and SUMO-1, and spacefill representation of their electrostatic surface potentials. **A.** Overlay of Ub (red) and SUMO-1 (black) in back-bone representations shows the similarity between the two protein cores, and the unique extended N-terminus of SUMO-1. Lys48 of Ub which is known to be involved in polyubiquitin conjugate formation, is not conserved in SUMO-1 (Gln69). Diglycine motifs are indicated on the flexible C-termini of Ub and SUMO-1, and are in similar positions relative to the overall protein folds, for both proteins (taken from (Bayer *et al.*, 1998)). **B.** Spacefill representations of the electrostatic surface potentials of Ub and SUMO-1 (1-97) as predicted by Swiss Pdb Viewer v3.6 b2 (Guex *et al.*, 1999). Electrostatic calculations are based upon a dielectric constants of 80.0 (solvent) and 4.00 (protein) and physiological pH. The predominantly negative (shown in red) surface potential of SUMO-1 in comparison to the more neutral and basic (shown in white and blue respectively) regions in Ub, can clearly be seen.

A


Gly76
Gly75
Gly96
Gly97
Gln69
Lys48

B


Ubiquitin (1-76)
Gly76
SUMO-1 (1-97)
Gly 97

Furthermore, ΔN21-SUMO-1 actually appeared to exclude endogenous WT-SUMO-1 itself from the PML NBs. An explanation for this, concurrent with the theory that the charged N-terminal tail may play a role in SUMO-1 specific protein-protein interactions, is that it may be involved in SUMO-1 protease recognition. If the *in vivo* dynamic regulation of ΔN21-SUMO-1 favoured conjugation over removal, then normal cellular turnover would eventually result in a smaller pool of free ΔN21-SUMO-1, and a greater number of ΔN21-SUMO-1 conjugated proteins than those modified by endogenous WT-SUMO-1. This theory could be fairly simply investigated using *in vitro* protease assays.

Further theories for the role of these charged regions in SUMO-1 may be that they are involved in the functional or structural modifications of substrate proteins. Considering that only SUMO-1-modified RanGAP1 interacts with RanBP2, and neither SUMO-1 nor RanGAP1 alone contain sufficient targeting information, SUMO-1-modification of RanGAP1 may involve a conformational change in the protein. Thus it is plausible that the structural and electrostatic charge differences between SUMO-1 and Ub may be involved in the functional or structural consequence of modification to target proteins.

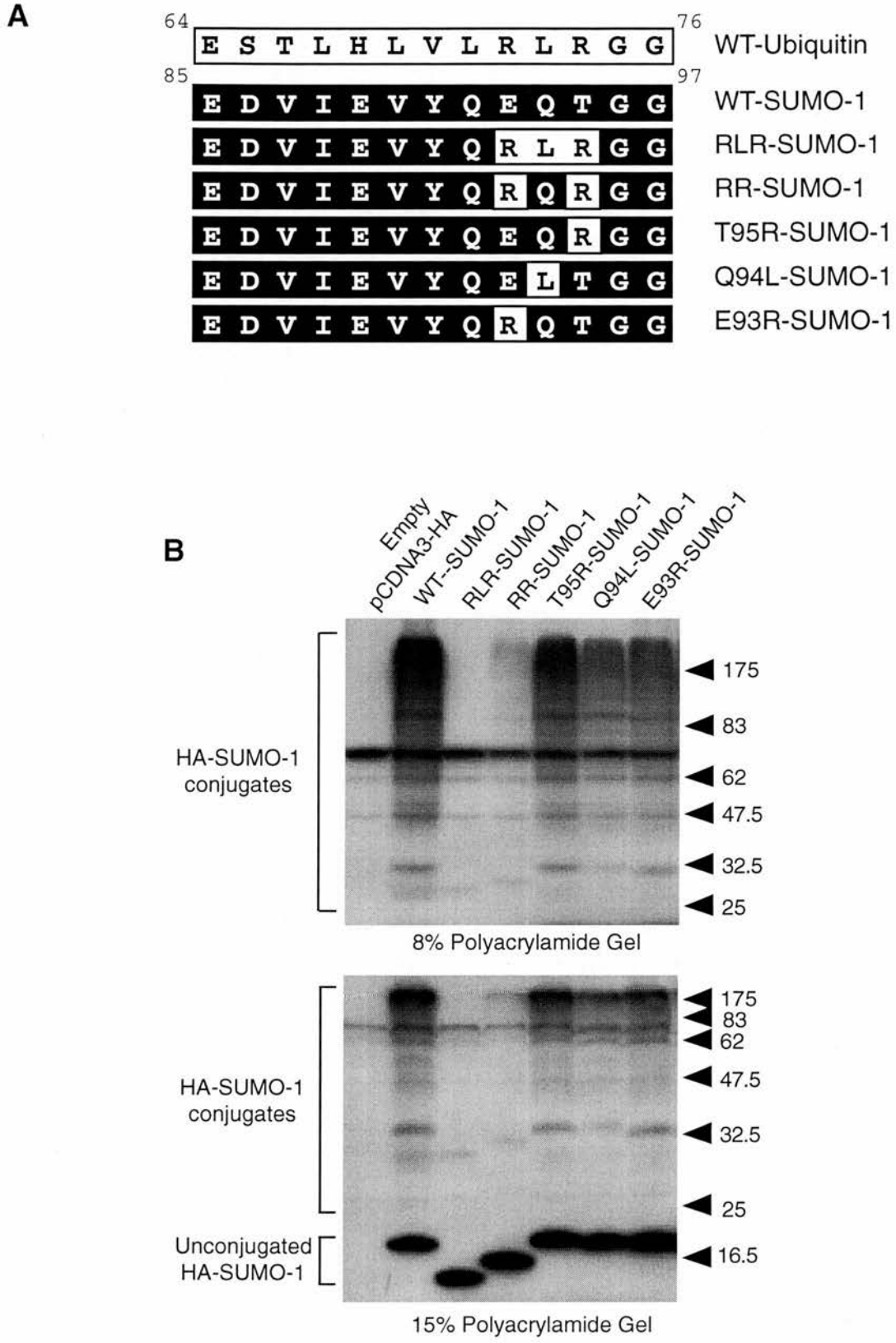
3.2. Role of the SUMO-1 C-terminus in Enzymatic Linkage to Protein Substrates

3.2.1. Results

3.2.1.1. Amino-acids adjacent to the C-terminal diglycine are important for SUMO-1 modification in vivo

Proximal to the C-terminal diglycine in SUMO-1 is a negatively charged region that is positively charged in the corresponding region of Ub (Figure 8B). To investigate the role of this region in discrimination by the enzymes of the two respective conjugation pathways, SUMO-$1_{(1-97)}$ molecules were generated containing various substitutions of amino-acids Glu93, Gln94 and Thr95 with Arg72, Leu73 and Arg74 from Ub (Figure 9A). The double and triple mutants (RR-SUMO-1 and RLR-SUMO-1) involve the largest net charge difference in that region of +3, while E93R-SUMO-1 and T95R-SUMO-1 involve net charge changes of +2 and +1 respectively. The mutation Q94L-SUMO-1 generates no net charge change. To demonstrate their conjugation efficiency *in vivo* the mutant and WT HA-SUMO-1-pcDNA-3 constructs were transiently transfected in COS7 cells and SDS lysates analysed by Western-blotting (Figure 9B upper and lower panels). It should be noted that the higher electrophoretic mobilities of the mutants RLR-SUMO-1 and RR-SUMO-1 as seen in the 15% acrylamide gel, are due to the charge differences imposed by the mutations, and do not represent cleaved protein products. This was confirmed by mass

Figure 9. Amino-acids adjacent to the C-terminal diglycine are important for SUMO-1 modification *in vivo.* **A.** C-terminal mutants of SUMO-1 were designed such that amino acids 93 to 95 were substituted with their equivalent counterparts from Ub relative to the C-terminal diglycine. **B.** COS7 cells were transfected as indicated with WT-SUMO-1 or mutant DNA constructs in pCDNA3 plasmid with an N-terminal HA tag. After 36 hours of expression SDS lysates were prepared, and exogenous expression was analysed by Western-blotting with the HA-Tag specific mAb 12CA5. Unconjugated SUMO-1 WT/mutants were resolved by electrophoresis in 15% polyacrylamide gels containing SDS (lower panel), and conjugated species by electrophoresis in 8% polyacrylamide gels containing SDS (upper panel). Molecular weight markers and the positions of conjugated and unconjugated SUMO-1 species are indicated


A
64
E S T L H L V L R L R G G
76
WT-Ubiquitin
85
E D V I E V Y Q E Q T G G
97
WT-SUMO-1
E D V I E V Y Q R L R G G
RLR-SUMO-1
E D V I E V Y Q R Q R G G
RR-SUMO-1
E D V I E V Y Q E Q R G G
T95R-SUMO-1
E D V I E V Y Q E L T G G
Q94L-SUMO-1
E D V I E V Y Q R Q T G G
E93R-SUMO-1
B
Empty pCDNA3-HA
WT-SUMO-1
RLR-SUMO-1
RR-SUMO-1
T95R-SUMO-1
Q94L-SUMO-1
E93R-SUMO-1
HA-SUMO-1 conjugates
175
83
62
47.5
32.5
25
8% Polyacrylamide Gel
HA-SUMO-1 conjugates
Unconjugated HA-SUMO-1
175
83
62
47.5
32.5
25
16.5
15% Polyacrylamide Gel

spectrometry of recombinant proteins, which displayed an indistinguishable pattern of electrophoretic mobilities (Figure 11A). Inspection of the 12-20kDa region of the 15% acrylamide gel reveals that unconjugated SUMO-1 proteins were produced at equivalent levels. As expected, the accumulation of conjugated species, seen more clearly on the 8% polyacrylamide gel, was reduced with respect to the wild type in cells transfected with each of the mutants. As shown previously HA-WT-SUMO-1 is conjugated to multiple endogenous polypeptides (Desterro *et al.*, 1998), whereas HA-RLR-SUMO-1, the mutant with the same 5 C-terminal amino-acids as Ub, appears not to form any high molecular weight adducts. Similarly the double mutation HA-RR-SUMO-1 appears to be almost entirely defective in its ability to form adducts with endogenous targets. Interestingly, the single mutations reveal that the variant Gln94 to Leu (HA-Q94L-SUMO-1) has the greatest single effect on conjugation. The mutant Glu93 to Arg (HA-E93R-SUMO-1) appears to have a higher conjugation efficiency than Q94L-SUMO-1, while the mutation of Thr95 to Arg (T95R-SUMO-1) is almost as active as the WT.

3.2.1.2. Role of the C-terminal amino-acids in subcellular localisation of SUMO-1

To analyse the sub-cellular localisation of the C-terminal SUMO-1 mutants HeLa cells were transfected with the HA-tag pcDNA-3 constructs described previously. Localisation of species reactive to anti-HA and anti-PML antibodies were detected using indirect immunofluorescence

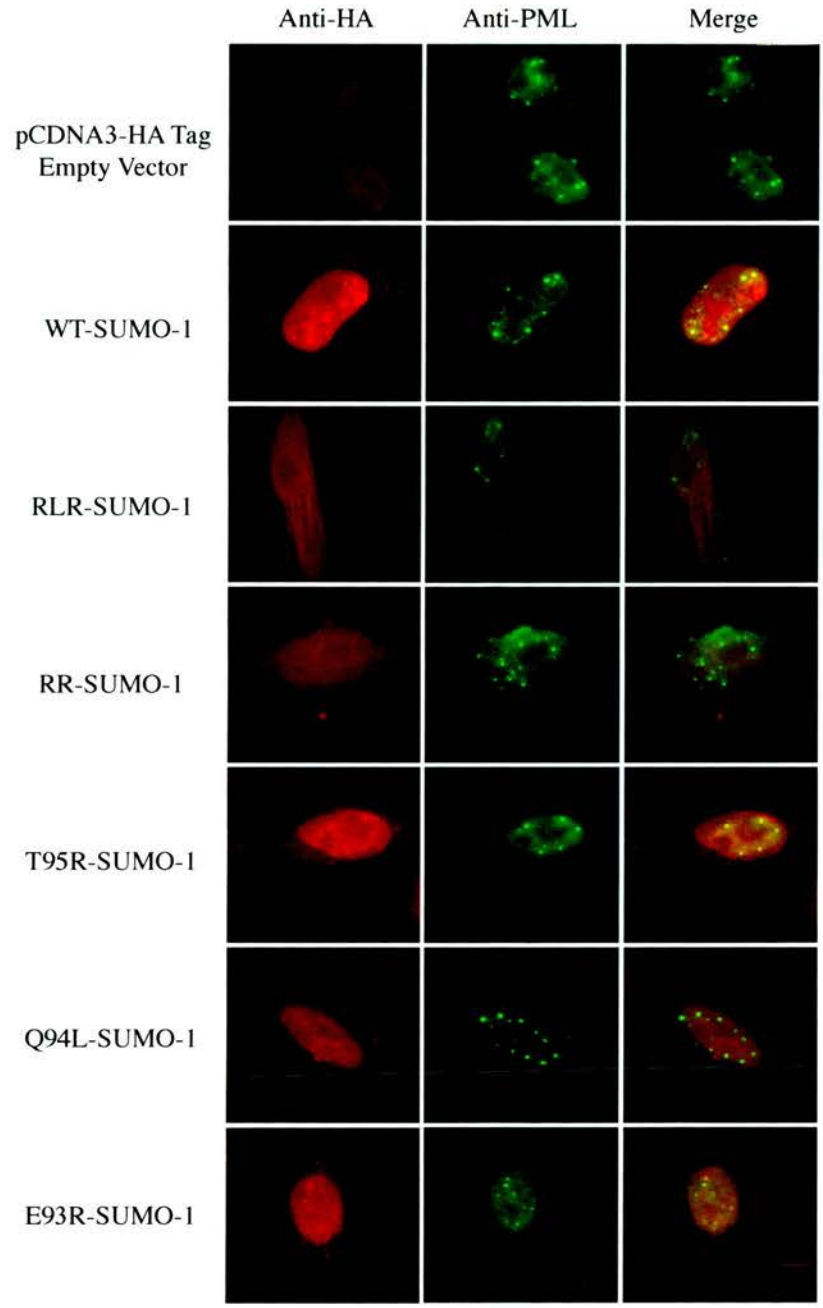


Figure 10. Role of the C-terminal amino-acids in subcellular localisation of SUMO-1. HeLa cells were electroporated with the WT and C-terminal mutants of SUMO-1 in HA-tag pCDNA3, or with empty vector alone, and processed for immunofluorescence as detailed in the Materials and Methods section. Slides were analysed by fluorescence microscopy at 100x magnification under oil immersion. Exogenous HA-tagged proteins were detected with (Babco) poly-HA and endogenous PML by a mAb 5E10. Texas-Red anti-rabbit and FITC anti-mouse were used as secondary antibodies. Thus HA-tagged proteins are shown in red and PML in green. Image exposure times and subsequent manipulations were constant for each fluorescence wavelength.

microscopy (Figure 10). As seen previously WT-SUMO-1 is found predominantly in the nucleus of cells and co-localises with endogenous PML in nuclear bodies (Boddy *et al.*, 1996; Stuurman *et al.*, 1992). All the single residue change mutants HA-T95R-SUMO-1, HA-Q94L-SUMO-1 and HA-E93R-SUMO-1 maintain their ability to localise in the nucleus and accumulate in PML nuclear bodies, whereas the double and triple amino-acid change mutants, HA-RR-SUMO-1 and HA-RLR-SUMO-1, are diffusely distributed throughout the cell, and do not accumulate in PML nuclear bodies. The inability of HA-RLR-SUMO-1 and HA-RR-SUMO-1 to accumulate in the nucleus is consistent with their inability to form high molecular weight conjugates in COS7, suggesting that the fluorescence pattern seen is predominantly due to unconjugated molecules. These results are consistent with previously published data regarding an inactive SUMO-1 mutant lacking both of the C-terminal diglycine residues (SUMO-1ΔC6). This mutant was found to be completely incapable of forming conjugates to substrates *in vivo* and *in vitro* and can be found distributed diffusely throughout COS7 cells (Mahajan *et al.*, 1998).

3.2.1.3. Role of the C-terminal amino-acids in SUMO-1 conjugation *in vitro*

To analyse the efficiency by which the C-terminal SUMO-1 mutants are conjugated to specific substrates an *in vitro* assay system was employed (Desterro *et al.*, 1998). PML and p53 were used as substrates as both have been shown previously to be covalently modified by SUMO-1

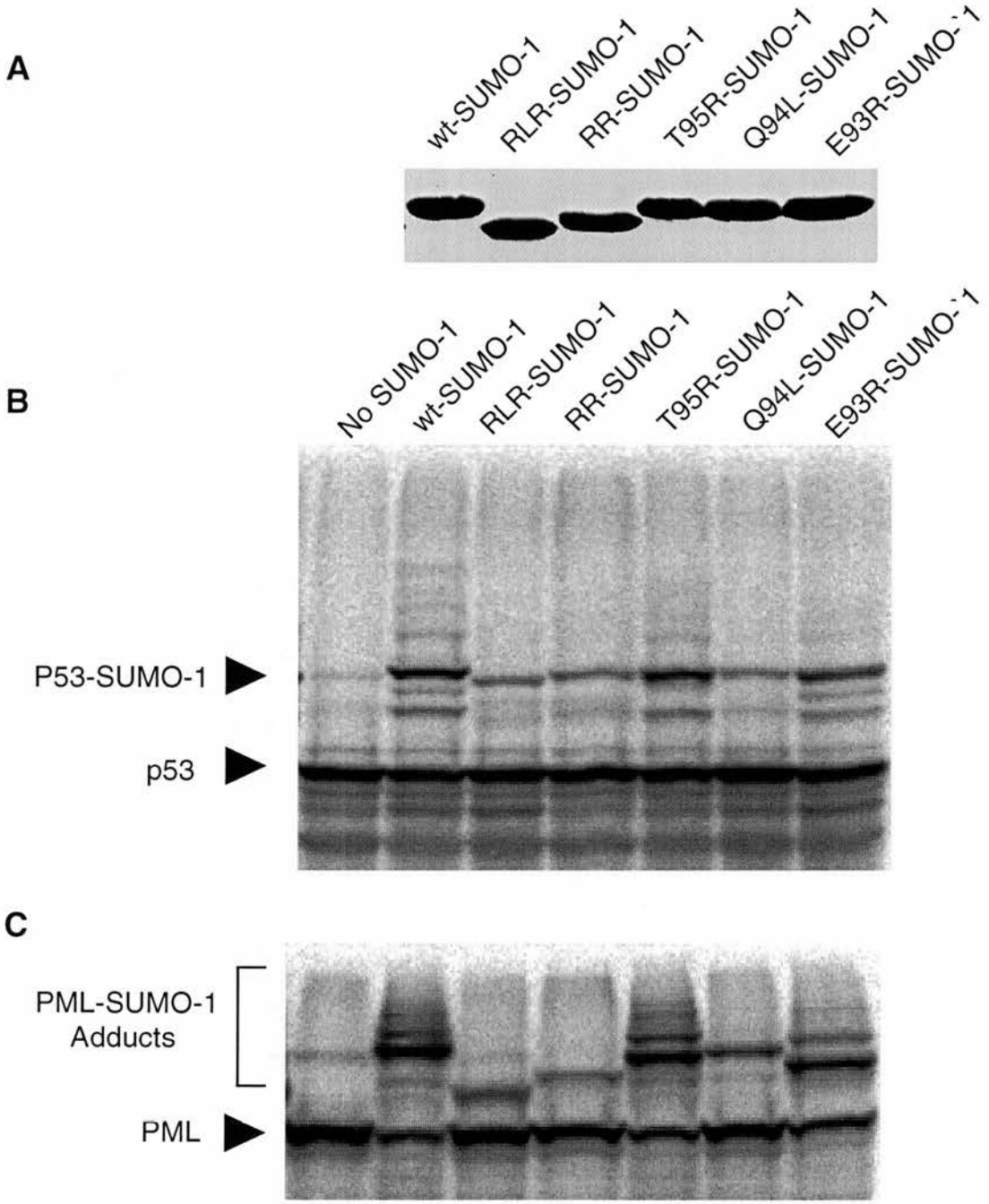


Figure 11. Role of the C-terminal amino-acids in SUMO-1 conjugation *in vitro*. A. 10µg purified recombinant SUMO-1 molecules were separated by electrophoresis in a 15% polyacrylamide gel containing SDS before Coomassie staining, destaining, and photography. ^{35}S-labelled, *in vitro* translated p53 (B) and PML (C) were incubated in the assay for SUMO-1 conjugation with equal quantities of WT and mutant SUMO-1 proteins as indicated. Reaction products were fractionated by SDS-PAGE before the gels were dried and analysed by phosphorimaging. The positions of the free and conjugated forms of p53 and PML are indicated.

(Duprez *et al.*, 1999; Kamitani *et al.*, 1997b; Muller *et al.*, 1998a; Rodriguez *et al.*, 1999; Sternsdorf *et al.*, 1997). SUMO-1 proteins were expressed in *E. coli.* and the purified proteins analysed by Coomassie stained SDS PAGE (Figure 11A) and MALDI TOFF mass spectrometry. All the proteins were of the predicted molecular weight and displayed the same pattern of electrophoretic mobilities observed when the corresponding cDNAs were over-expressed in eukaryotic cells (Figure 9B). PML and p53 were labelled with ^{35}S-methionine *in vitro* and used separately as substrates in the assay with equal quantities of the WT-SUMO-1 or mutants. The efficiency by which the SUMO-1 mutants are conjugated to both p53 and PML is reduced to varying degrees compared to that of the WT protein (Figure 11 B and C). Comparisons with the COS7 cell data reveals that the degree of this reduction in activity is similar *in vivo* and *in vitro*. Quantitation of conjugated species by phosphorimaging demonstrates that for both substrates RLR-SUMO-1 only retains approximately 18% of the WT activity. RR-SUMO-1 and Q94L-SUMO-1 are only marginally more active with ~24% and ~21% of WT activity respectively. E93R-SUMO-1 however, appears to retain approximately 63% activity while T95R-SUMO-1 is almost as efficiently conjugated as the WT protein with approximately 93% activity.

3.2.1.4. Role of the SUMO-1 C-terminus in substrate selection by the E1 and E2 enzymes

SUMO-1 is conjugated to target substrates via two specific enzyme activities. In the first step the C-terminal glycine of SUMO-1 forms a thioester bond with the putative active-site cysteine 173 in the SAE2 subunit of the SUMO activating enzyme (E1). This high energy thioester is then transferred to cysteine 93 in Ubc9 which acts as the SUMO-1 conjugating enzyme (E2). Finally the C-terminal glycine of SUMO-1 forms an isopeptide bond with a specific lysine in the target protein. To identify the level within this pathway at which the C-terminal SUMO-1 mutants are defective a SUMO-1 thioester assay was employed (Desterro *et al.*, 1997). The SUMO-1 WT and mutant proteins were labelled with ^{125}Iodine and iodoacetamide treated to limit the formation of SUMO-1~SUMO-1 disulfide dimers via the internal cysteine 52. Equal quantities of each mutant and the WT were incubated in the assay with and without Ubc9 as described. In polyacrylamide gels under non-reducing conditions the W T SUMO-1~E1 thioester can be seen in the absence of Ubc9, and the SUMO-1~E2 can be seen in its presence (Figure 12). These radioactive species are eliminated by incubation under reducing conditions, confirming that they are thioester linked proteins (Data not shown). Neither RLR-SUMO-1 nor RR-SUMO-1 form a thioester bond with the SUMO-1 E1, which as a result cannot form a thioester with Ubc9. Similarly E93R-SUMO-1 inefficiently forms an E1 thioester, suggesting Glu93 is involved in some part in the recognition of SUMO-1 by the E1 activating enzyme. In contrast, T95R-SUMO-1 which appears to retain almost WT activity levels in overall

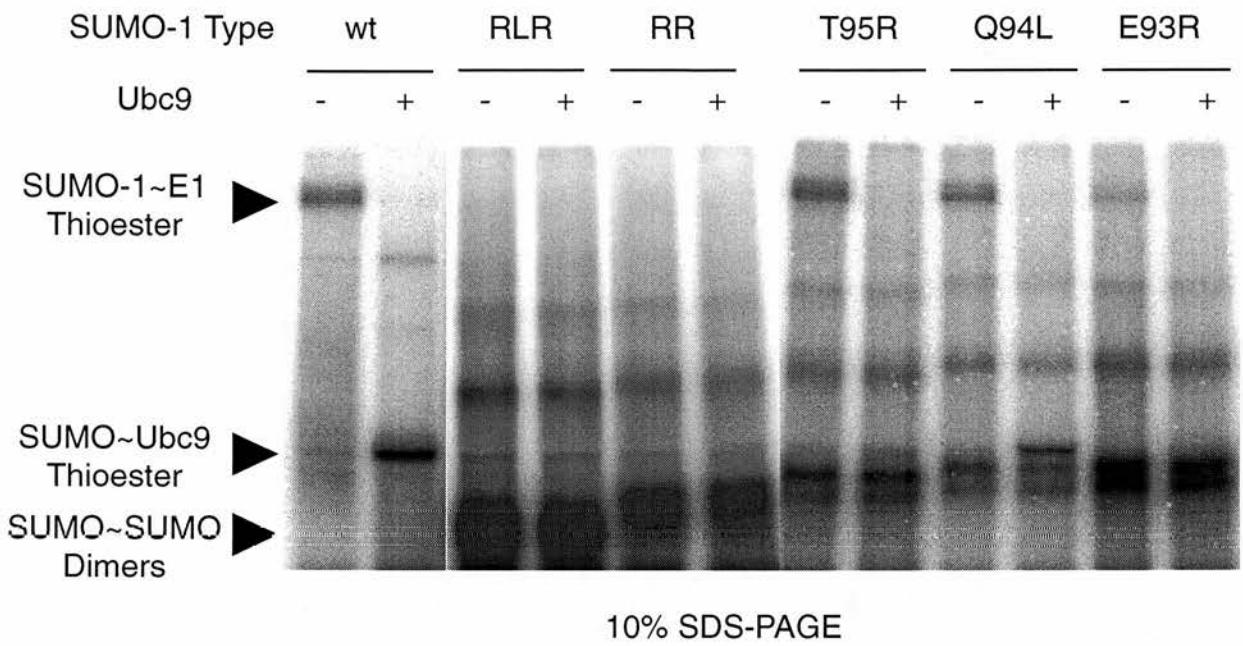


Figure 12. Role of the SUMO-1 C-terminus in substrate selection by E1 and E2 enzymes. Equal quantities of [^{125}I] labelled WT and mutant SUMO-1 proteins were incubated in the assay for thioester formation in both the presence and absence of 1μg Ubc9 as indicated. After 30 minutes at 30°C, reactions were terminated and products fractionated by electrophoresis in 10% polyacrylamide gels containing SDS under non-reducing conditions. Gels were dried and ^{125}I incorporation determined by phosphorimaging. The positions of the ^{125}I SUMO-1 thioester adducts with SUMO E1 and Ubc9 are indicated.

SUMO-1 Type	Net Charge Difference From wt	Approx. % of wt *in vitro* Activity	Level of Effect in Conjugation Pathway
Wild Type	-	100	-
RLR	+3	18	E1
RR	+3	24	E1
T95R	+2	93	E2
Q94L	0	21	Transfer to Substrate
E93R	+1	62	E1

Table 5. Summary of the effects of C-terminal SUMO-1 mutants in the conjugationpathway. Summary of the net charge change with respect to Ub of the C-terminal mutations, their effect on overall conjugation efficiency *in vitro*, and the level within the SUMO-1 conjugation pathway at which they have their effect.

conjugation efficiency and E1 thioester formation, does not form a Ubc9 thioester as readily as the WT. This indicates the importance of Arg95 in the efficient interaction between the two proteins. Q94L-SUMO-1, which is the least active of all the single mutations in overall conjugation efficiency, retains the ability to form both E1 and E2 thioesters. This suggests that the substitution of the hydrophilic glutamine of SUMO-1 for the hydrophobic leucine of Ub inhibits the transfer of SUMO-1 onto target substrates.

3.2.2. Discussion

SUMO-1 is conjugated to specific lysine residues on target proteins by an enzymatic pathway involving SAE1/SAE2, and Ubc9. Although the conjugation mechanisms of Ub and SUMO-1 are similar, they differ in how they appear to achieve specificity. The SAE1/SAE2 heterodimer activates SUMO-1 and not Ub, but shares a number of domains with the monomeric Ub E1 UBA1 (Figure 4), including the region containing the active site cysteine residue (Desterro *et al.*, 1999). Protein precursors of both SUMO-1 and Ub are cleaved by specific C-terminal hydrolases to reveal a diglycine motif, which results in a 76 amino-acid molecule for Ub, and one of 97 residues for SUMO-1 (Hershko and Ciechanover, 1998; Li and Hochstrasser, 1999). SUMO-1 and Ub share a very high degree of structural similarity, although they differ markedly in their surface charge potential and distribution, with SUMO-1 presenting a number of negatively charged regions which are either uncharged or positively charged in Ub (Bayer *et al.*, 1998). It is these regions which are proposed to be involved in the discrimination between the enzymes of the two conjugation pathways. One such region is the C-terminal 5 residues of the two protein modifiers (Figure 9A). The terminal glycine 97 involved in E1 and E2 thioesters and in the isopeptide link to the target, is adjacent to a negatively charged region in SUMO-1 and a positively charged region in Ub. The data presented here demonstrate that the substitution of amino-acids 93-95 from SUMO-1 for the respective residues from Ub (See Figure 9A)

generates SUMO-1 molecules which vary in their abilities to form thioester bonds with the SUMO E1 and E2 enzymes, and to be transferred from E2 to target substrates (summarised in Table 5).

Previous studies of both Ub and the UBL NEDD8 show that the amino-acid fifth from the C-terminal Glycine (Arg72 of Ub and Ala72 of NEDD8) are involved in E1 binding (Burch and Haas, 1994; Whitby *et al.*, 1998). Consistent with this data, our studies show that the mutation of the equivalent residue in SUMO-1 (Glu93) which constitutes a net charge change of +2, significantly reduces the ability of the protein to form a thioester bond with the SUMO-1-specific E1. The *in vitro* conjugation efficiency of this mutant is approximately 63% of the WT, which is consistent with the *in vivo* data showing a lower overall conjugation efficiency to endogenous substrates in COS7 cells than the WT SUMO-1. While the E93R mutation is also is predominantly nuclear, the degree of co-localisation in with PML in nuclear bodies appears to be less than the WT.

Arginines 72 and 74 of Ub both contribute to the overall positive charge of the C-terminus of Ub. Although Arg74 has been shown to be involved in E1 binding for Ub (Duerksen-Hughes *et al.*, 1987), our data shows that the mutation of Thr95 (equivalent to Arg74) to Arg in SUMO-1 (T95R-SUMO-1) appears to have very little effect on the overall conjugation efficiency of the protein both *in vivo* and *in vitro*. It retains approximately 93% of the WT conjugation efficiency *in vitro*, and the sub-cellular localisation of this mutant in HeLa cells is almost indistinguishable from that of the WT. Although this mutation has no

effect on E1 thioester formation, the ability of the protein to form a Ubc9 thioester is significantly reduced. Since this mutant is almost as readily conjugated to substrates as WT SUMO-1, this demonstrates that the transfer of SUMO-1 from the E1 to Ubc9 is not the rate limiting step in the conjugation pathway *in vitro* or *in vivo*. It is also worth noting that although the efficiency of the transfer reaction from E1 to E2 is reduced for this mutant, this does not result in the accumulation of the E1~SUMO-1 thioester intermediate. One explanation for this observation is that even although the E1 cannot efficiently transfer T95R-SUMO-1 to Ubc9, the presence of bound Ubc9 may enhance the rate of the reverse reaction, resulting in the release of free T95R SUMO-1.

The conjugation of Ub to substrates is known to require not only E1 and E2 activities, but a third enzyme known as a Ub ligase (E3). The Ub E3 is thought to act as the recognition protein in the pathway, with specific E3s being responsible for the targeting of the E2~Ub complex to specific substrates. Although conjugation of SUMO-1 to protein targets *in vitro* does not require the presence of an E3 activity, it remains a possibility that such activities may enhance the rate of conjugation *in vivo*. However, in contrast to Ub conjugation, the lysine residues to which SUMO-1 is conjugated all appear to exist within a particular motif, ΨKXE (where ψ represents I, F, V or L, and x is any amino-acid) (Rodriguez *et al.*, 1999; Sternsdorf *et al.*, 1999). This motif may act as the recognition site for Ubc9 or Ubc9~SUMO-1, and thus provide specificity in SUMO-1 modification. Q94L-SUMO-1 is not efficiently conjugated to protein substrates *in vivo* or *in vitro* although its ability to form thioester bonds with the E1 and E2

Figure 13. 3D electrostatic surface potential representations of WT-SUMO-$1_{(1-97)}$, the mutant RLR-SUMO-1 and $Ub_{(1-76)}$. Spacefill representations of electrostatic potential maps of WT-SUMO-1, the RLR mutant, and Ub as predicted by Swiss Pdb Viewer v3.6 b2 (Guex *et al.*, 1999). Electrostatic calculations are based upon a dielectric constants of 80.0 (solvent) and 4.00 (protein) and physiological pH. Regions of net negative charge are shown in red, and basic shown in blue. Two projections showing N- to C-terminal "Side" views and an "End" view showing the regions surrounding the C-terminal glycines (Yellow).

WT-SUMO-1
GLY 97
RLR-SUMO-1
GLY 97
WT-Ubiquitin
GLY 76

enzymes is almost at WT levels. Thus substitution of the hydrophilic glutamine group for the hydrophobic leucine present in Ub inhibits the formation of the isopeptide bond with target lysine residues. It is therefore possible that the ΨKxE motif is recognised by elements from both Ubc9 and SUMO-1 that are brought together in the Ubc9~SUMO-1 thioester complex.

RLR-SUMO-1 is a very poor substrate for SUMO-1 conjugation and although it has the same 5 C-terminal amino-acids as Ub it cannot form a thioester with either the SUMO-1 or the Ub E1 (data not shown). Thus the C-termini of Ub and SUMO-1 are required, but are not sufficient for recognition by the respective E1 enzymes. Further mutagenesis will be required to fully characterise the elements in SUMO-1 and Ub required for E1 recognition.

3.3. Role of SUMO-1 C-terminus in Maturation and Deconjugation

3.3.1. Results

3.3.1.1. Single and double SUMO-1 C-terminal mutants appear to undergo efficient maturation *in vivo*

Like almost all members of the UBL family of proteins, SUMO-1 is primarily synthesised as an inactive precursor (SUMO-$1_{(1\text{-}101)}$), requiring cleavage by specific C-terminal hydrolase enzymes to remove the inhibitory peptide (maturation), producing the active protein (SUMO-$1_{(1\text{-}97)}$, also known as SUMO-1-GG). Studies involving SUMO-1-GG-C-terminal fusion proteins which are cleaved as efficiently as the WT protein with the His/Ser/Thr/Val (HSTV) tag, show that residues following Gly97 in SUMO-1 are not significant determinants on protease activity (Kim *et al.*, 2000; Li and Hochstrasser, 1999; Suzuki *et al.*, 1999). To investigate the significance of residues prior to the diglycine of SUMO-1 in C-terminal hydrolase activity, the mutants of SUMO-1 with conserved residues from Ub, shown to be important for conjugation (sections 3.3 and 3.4 of this work, see Figures 9A for sequences) were reconstructed with the HSTV peptide tag (Figure 17) using PCR. These SUMO-$1_{(1\text{-}101)}$ mutant DNA constructs were cloned into the transient transfection plasmid pcDNA-3 with an N-terminal HA tag. Five mutant DNAs were produced;

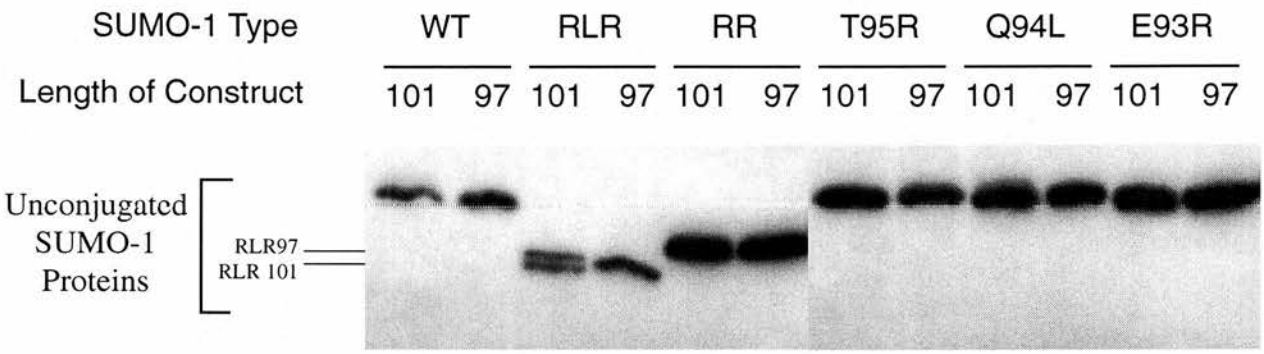


Figure 14. Only the triple residue C-terminal mutant of SUMO-1significantly affects *in vivo* C-terminal hydrolase activity. COS7 cells were transfected as indicated with WT-SUMO-1 or mutant DNA constructs in pCDNA3 plasmid with an N-terminal HA tag. After 36 hours of expression SDS lysates were prepared, and exogenous expression was analysed by Western-blotting with the HA-Tag specific mAb 12CA5. Unconjugated SUMO-1 WT/mutants were resolved by electrophoresis in 15% polyacrylamide gels containing SDS.

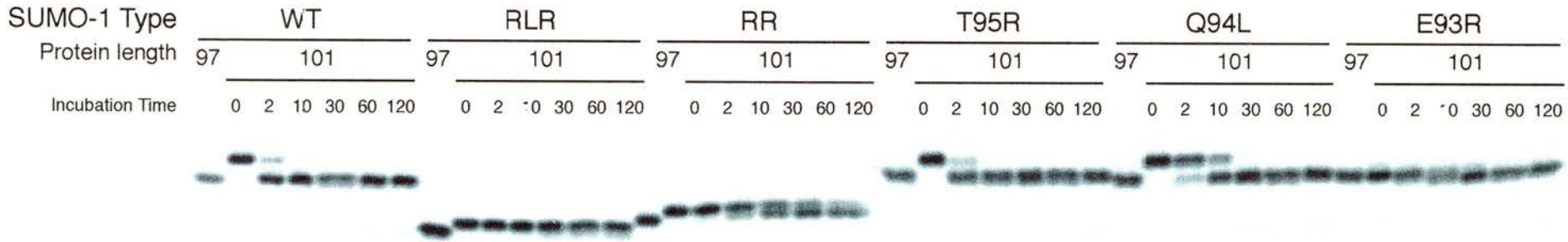
pcDNA-3-HA-RLR-SUMO-$1_{(1-101)}$, pcDNA-3-HA-RR-SUMO-$1_{(1-101)}$, pcDNA-3-HA-T95R-SUMO-$1_{(1-101)}$, pcDNA-3-HA-Q94L-SUMO-$1_{(1-101)}$ and pcDNA-3-HA-E93R-SUMO-$1_{(1-101)}$. To determine their susceptibility to *in vivo* C-terminal hydrolase activity, COS7 cells were transfected with pcDNA-3-HA plasmids containing, each 101 residue SUMO-1 mutant and the WT-SUMO-$1_{(1-101)}$, as well as their respective 97 residue equivalents, and cell lysates analysed by Western-blotting for HA reactive species (see Figure 14). With the exception of the triple mutant, HA-RLR-SUMO-$1_{(1-101)}$, which is present in transfected cells as both the 101 and 97 residue forms, all SUMO-1 mutant proteins appear to be fully cleaved 24h post-transfection. It should be noted that due to the fact that HA-E93R-SUMO-$1_{(1-101)}$ and HA-E93R-SUMO-$1_{(1-97)}$ may have the same apparent molecular weight (see *in vitro* translation products in Figure 15), it is impossible to conclude the extent to which the full-length construct is cleaved *in vivo*. However, as the pattern of high molecular weight conjugates was identical for both the 1-101 and 1-97 transfected cells (data not shown), it can be concluded that the full-length construct was at least partially cleaved by endogenous proteases to yield enough mature HA-E93R-SUMO-1 for conjugation to substrates.

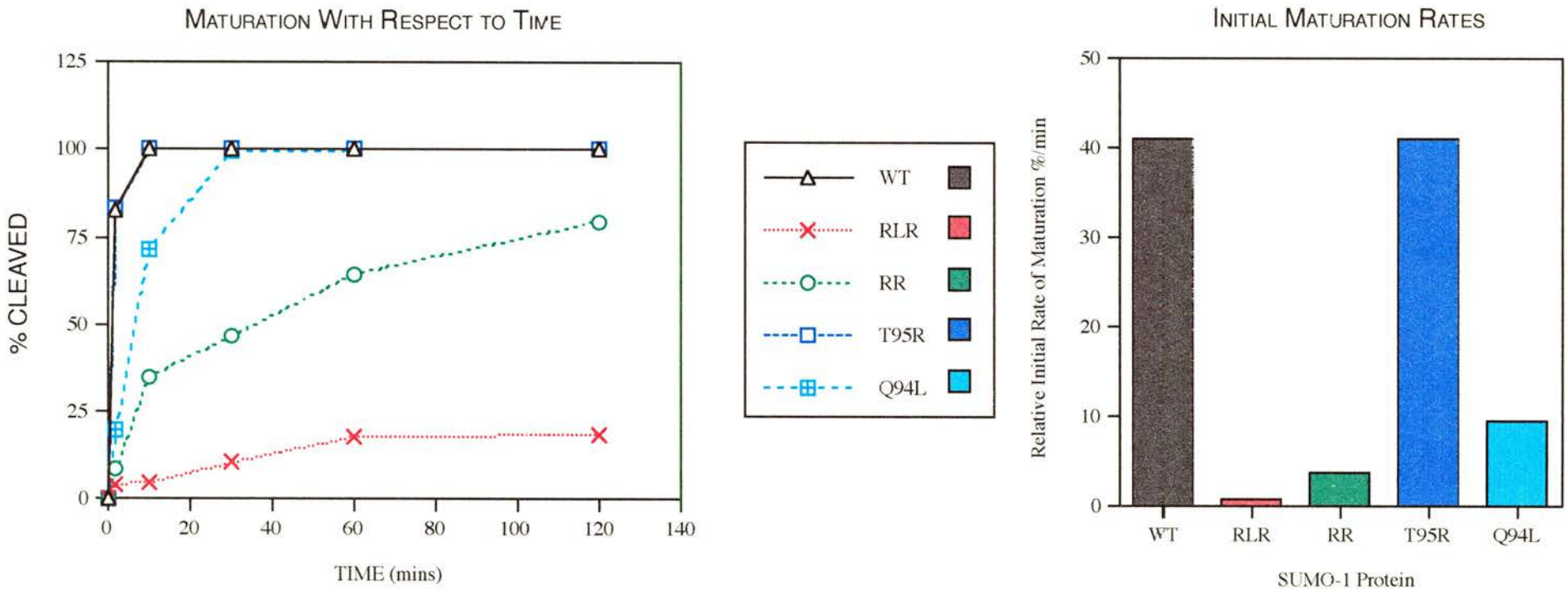
3.3.1.2. *In vitro* SUMO-1 maturation proceeds at varying rates for C-terminal mutants

To quantitatively investigate the effects of the C-terminal mutations of SUMO-1 on the ability of the protein to be processed by SUMO-1

hydrolases, a simple *in vitro* assay was employed. ^{35}S *in vitro* translated 1-101 forms of each HA-SUMO-1 mutant and the WT were incubated for varying lengths of time at 30°C in the presence of 2µl of a crude HeLa cell nuclear extract known to contain SUMO-1-specific hydrolase activity. Full-length, and cleaved forms could be fractionated by SDS PAGE and analysed by phosphorimaging in comparison with their respective *in vitro* translated ^{35}S 1-97 residue forms (see Figure 15 upper panel). Again, discrimination between the 1-101 and 1-97 forms of each SUMO-1 protein is possible, with the exception of HA-E93R-SUMO-$1_{(1-101)}$, which appears to either be cleaved within the *in vitro* translation mix, or has the same retention factor as the mature protein. Upper bands (1-101 proteins) and lower bands (1-97), for the other proteins were quantified using MacBas software (MacBas V 2.5), and the percentage of cleaved protein was plotted against incubation time, from which initial rate constants were calculated in % cleaved per min (see Figure 15 graphs left and right respectively). It can be seen that HA-WT-SUMO-$1_{(1-101)}$ is cleaved rapidly, and within 10 min no full-length protein remains. An essentially WT pattern can be observed for the HA-T95R-SUMO-1 mutant, which is also completely cleaved in under 10 min, while the mutation of Q94L appears to extend the life of SUMO-1 to approximately 30 min in the assay. The cleavage of the double mutation of Thr95 and Glu93 to Arg (HA-RR-SUMO-1) appears to be fairly severely impaired and was only approximately 75% cleaved after 2h incubation. Furthermore, in agreement with the *in vivo* analysis, the triple mutant (HA-RLR-SUMO-1) was the poorest substrate for the C-terminal hydrolase activity from the crude HeLa nuclear extract.

Figure 15. C-terminal-mutants of SUMO-1 are cleaved *in vitro* with varying degrees of efficiency in comparison with WT-SUMO-1. SUMO-1 proteins of length 1-101 and 1-97 were ^{35}S-labelled by in vitro transcription/translation. ^{35}S-1-101 proteins were incubated (as batch) with a HeLa cell nuclear extract known to contain the SUMO-specific protease activity, and samples taken at different points in time as indicated. Reactions were terminated by the addition of SDS/β-mercaptoethanol boiling mix, samples boiled and fractionated by 12% polyacrylamide gel electrophoresis containing SDS before drying and exposure to phosphorimage screening (above). Quantification was performed using MacBas software "Quant" option (MacBas V 2.5). The percentage of cleaved protein with respect to total ^{35}S-labelled protein was calculated and presented graphically (bottom left), and the initial rates of hydrolysis were calculated and shown as a bar graph (bottom right).


SUMO-1 Type
WT
RLR
RR
T95R
Q94L
E93R
Protein length
97
101
Incubation Time
0 2 10 30 60 120

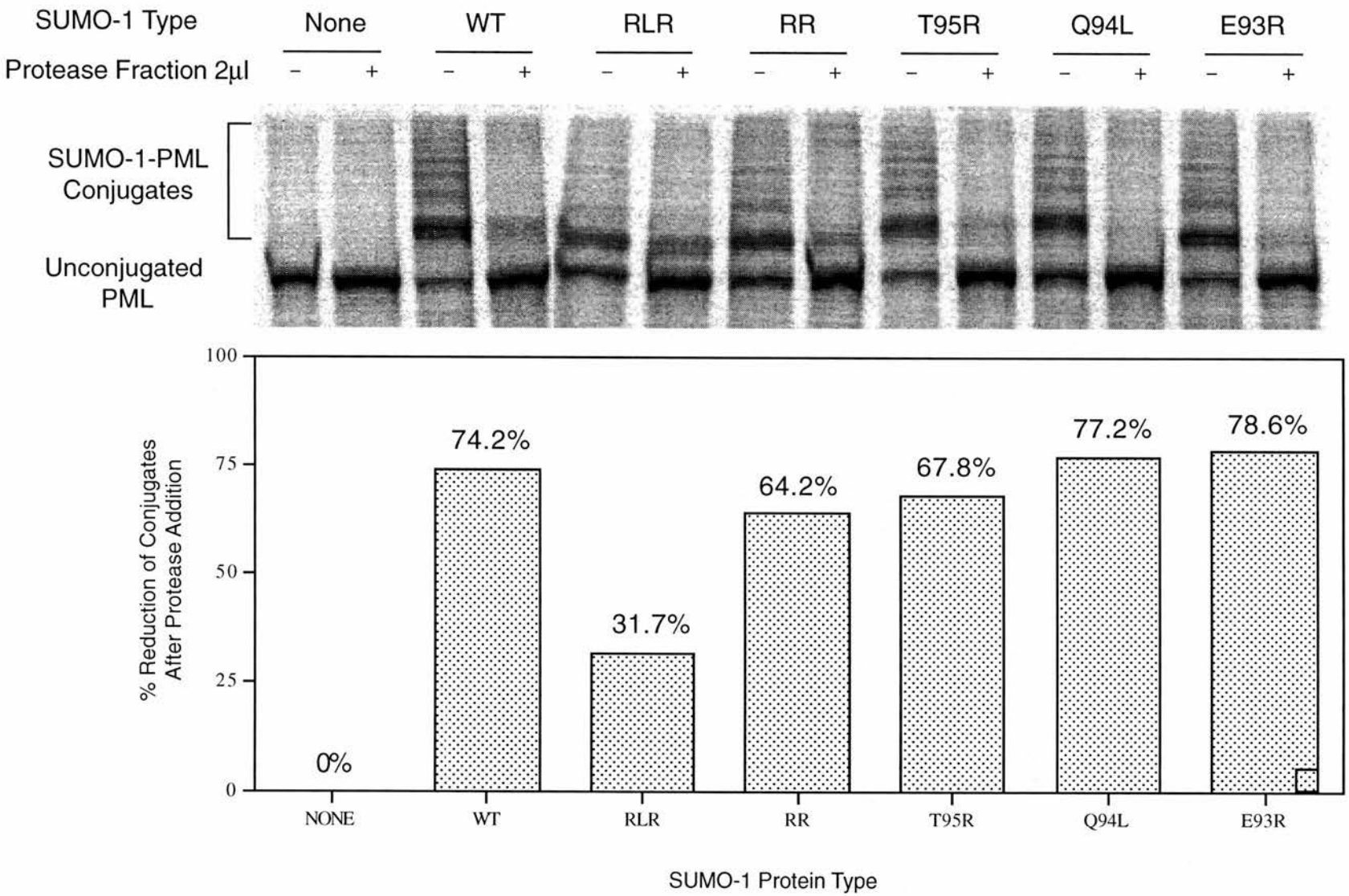

MATURATION WITH RESPECT TO TIME
% CLEAVED
TIME (mins)
0
25
50
75
100
125
0
20
40
60
80
100
120
140
WT
RLR
RR
T95R
Q94L
INITIAL MATURATION RATES
Relative Initial Rate of Maturation %/min
0
10
20
30
40
50
WT
RLR
RR
T95R
Q94L
SUMO-1 Protein

After 2h incubation, only 18% of the full-length protein was cleaved into the 97 residue form.

3.3.1.3. C-terminal mutants of SUMO-1 have a less significant effect on isopeptidase activity than maturation

To identify the effect the mutations of SUMO-1 have on the ability of the conjugated protein to be removed from substrate, the previously detailed SUMO-1 mutant conjugation assay was used to generate pools of conjugated protein, from which the SUMO-1 proteins could then be removed by the isopeptidase activity from the HeLa cell nuclear extract. The conjugation reactions used recombinant 1-97 SUMO-1 constructs and ^{35}S-PML as substrate. Reactions were incubated for 3h before termination with 10mM EDTA. A sample of reaction was taken before the addition of 2µl of the HeLa fraction, followed by further incubation at 30°C for an hour. Samples were then analysed by SDS PAGE and phosphorimaging (see Figure 16 upper panel). The quantities of SUMO-1-PML conjugated species were measured again using the 'Quant' function of the MacBas software, and the % by which these were reduced after the addition of HeLa fraction were calculated (see Figure 16 graph). The WT-SUMO-1-PML conjugates were reduced by approximately 75% after incubation with the isopeptidase-containing fraction for this period of time. In slight contrast with the C-terminal hydrolase results, T95R-SUMO-1 appears to be a marginally poorer substrate for the isopeptidase reaction than the WT protein (67.8%), while E93R-SUMO-1 and Q94L-SUMO-1 are removed to

Figure 16. C-terminal mutations of SUMO-1 marginally affect its removal from PML by the SUMO-1 protease *in vitro*. WT and C-terminal mutant proteins of SUMO-1 were incubated in the assay for conjugation to ^{35}S-labelled PML. The conjugation reaction was halted by addition of 10mM EDTA before addition (+) of the HeLa cell fraction containing the SUMO-1 protease activity and further incubation for 1h. Assays were fractionated by 8% Polyacrylamide gel electrophoresis before drying and exposing to phosphorimage screening (upper panel). The relative quantities of conjugated material both before and after the addition of protease were detected using MacBas Quantification (MacBas V 2.5), and represented graphically as a percentage reduction upon addition. Thus the lower the % reduction, the more resistant to deconjugation the SUMO-1 variant.


SUMO-1 Type
None
WT
RLR
RR
T95R
Q94L
E93R
Protease Fraction 2µl
- + - + - + - + - + - + - +
SUMO-1-PML Conjugates
Unconjugated PML
% Reduction of Conjugates After Protease Addition
100
75
50
25
0
0%
74.2%
31.7%
64.2%
67.8%
77.2%
78.6%
NONE
WT
RLR
RR
T95R
Q94L
E93R
SUMO-1 Protein Type

approximately the same extent as WT-SUMO-1 (78.6% and 77.2% respectively). Compared with WT-SUMO-1, the double mutant, RR-SUMO-1 is not as severely compromised in the isopeptidase activity, as it was for the hydrolase reaction. Again, the triple SUMO-1 mutant had the greatest effect on protease activity, as the quantities of RLR-SUMO-1-PML conjugates were reduced by less than half those of WT-SUMO-1 (only ~32%). This can be clearly seen in absolute terms in the crude data (Figure 16 upper panel). Even though before the addition of protease, the amounts of SUMO-1-PML conjugates are much lower for the RLR protein than the WT protein, after addition they are much higher.

It should be noted that the conjugation abilities of the SUMO-1 mutants in the experiment described here (Fig. 16) appear to differ from those shown in Figure 11, although the assay conditions are almost identical. The only difference between the two experiments was the stock of the ^{35}S-PML substrate used in the reactions. The batches of *in vitro* translated protein contain crude wheat-germ cell extracts of unknown, and possibly varying content. Thus it is possible that an unidentified protein from these extracts may be interfering with the assay by altering the specificity of the SUMO conjugation machenery. Although in itself this may be an interesting observation for future investigations, with respect to these studies, this highlights the requirement for an entirely characterised assay for conclusive analysis of such proteins.

3.3.2. Discussion

Mammalian cells contain a number of proteins with the ability to remove the HSTV tag from the C-terminus of the SUMO-1 primary translation product, as well as being able to remove SUMO-1 conjugated to target proteins (Suzuki *et al.*, 1999; Yeh *et al.*, 2000). Although they share no sequence or structure similarity, the roles of these SUMO-1-specific proteases can be regarded as analogous to those specific to Ub. The so-called sentrin proteases (SENPs) are a family of (to date) seven sequence related proteins thought to possess differing substrate specificity and cellular localisation (Yeh *et al.*, 2000). In an attempt to identify regions of SUMO-1 itself required for efficient processing and deconjugation, C-terminal mutants of the protein known to affect *in vitro* and *in vivo* conjugation were analysed for their ability to undergo maturation *in vivo* and *in vitro*, and to be removed from PML *in vitro*. Over-expression of the full-length SUMO-1 in COS7 cells showed that only the SUMO-1 mutant with three mutated residues in positions 93, 94 and 95, showed signs of accumulated full-length protein 24h after transfection. All other constructs, with the exception of the unresolvable HA-E93R-SUMO-1 mutant, could be seen to be efficiently cleaved in this time period. By using a crude HeLa cell fraction containing SUMO-1 C-terminal hydrolase activity/activities, this maturation process could be quantified with respect to time. Initial rates of the cleavage of full-length ^{35}S-labelled SUMO-1 proteins, show that the single mutation of Thr95 to Arg has no affect on the hydrolase activity,

giving an essentially WT profile of cleavage with respect to time. Interestingly, alignments of Smt3 homologues in the C-terminal region (Figure 17) show that of the three residues immediately prior to the diglycine in SUMO-1 mutated in this study, both Glu93 and Gln94 are conserved in both budding and fission yeast proteins, while Thr95 is not. An explanation for these observations is revealed by the recently resolved crystal structure of a deacylation intermediate complex between yeast Smt3 and an active C-terminal fragment of its specific protease, Ulp1(403-621) (Mossessova and Lima, 2000). Ulp1 is distantly related to known SUMO-1-specific proteases, and has dual specificity for both SUMO-1 and Smt3 (Li and Hochstrasser, 1999). Notably for this study, Glu94, Gln95 and Ile96 from Smt3 (equivalent to Glu93, Gln94 and Thr95 in SUMO-1), all make contacts with Ulp1, although Ile96 only interacts with Ulp1 via main-chain atoms (Mossessova and Lima, 2000). Clearly, the side-chain characteristics of residue 96 in SMT3 and 95 in SUMO-1 are unlikely to have any functional consequences, and as such the fact that a SUMO-1 Thr95 to Arg mutation has no effect on C-terminal hydrolase activity, is not surprising.

In contrast to Thr95 mutants, substitution of the invariant Gln94 for Leu produces a SUMO-1 protein with an initial hydrolysis rate four times slower than the WT. The likely reason for this partial stabilization is also revealed in the crystal structure of Smt3-Ulp1, which shows that unlike Ile96, the side-chain of Gln95 is involved in a number of significant interactions with Ulp1. Gln95 makes hydrogen bonds with Ser473 and Thr477 from Ulp1, as well as forming a pocket with Leu66 and His92, in

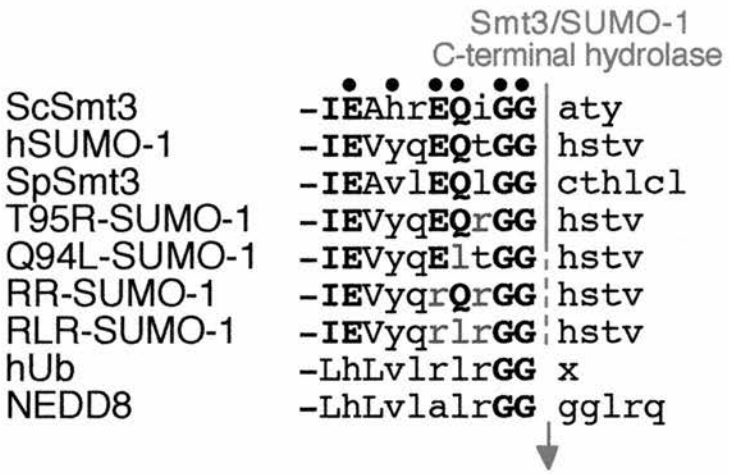
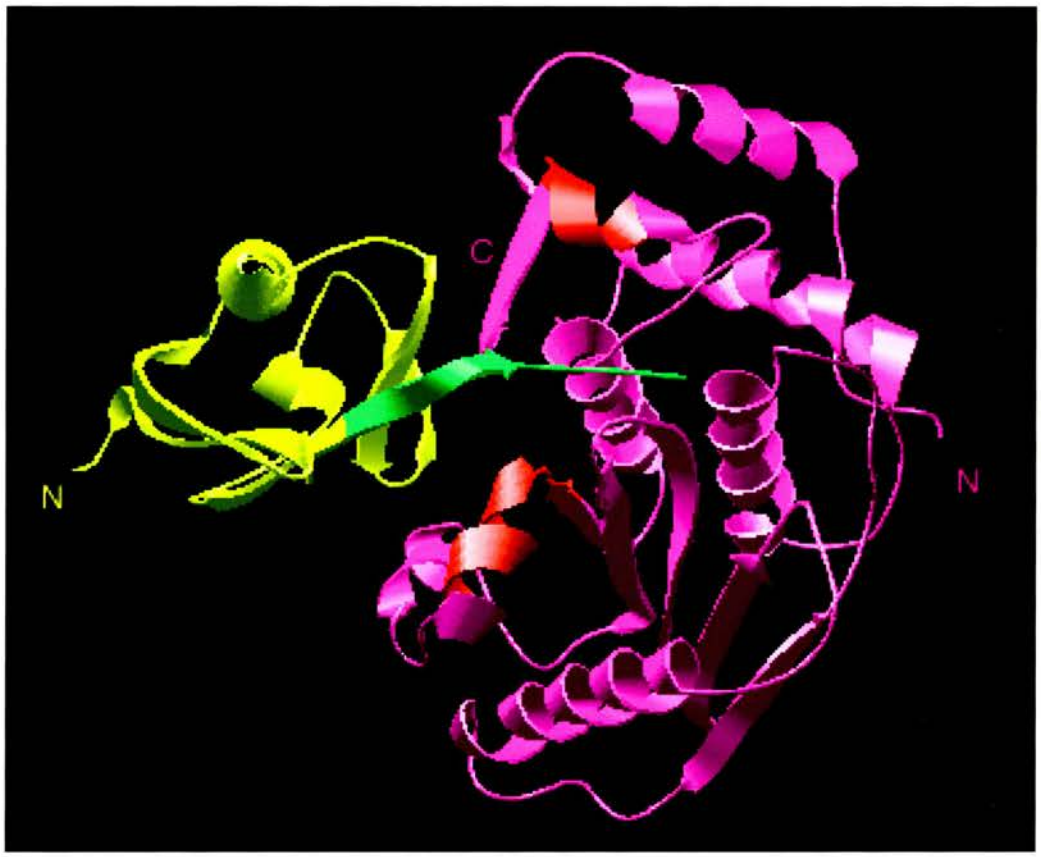


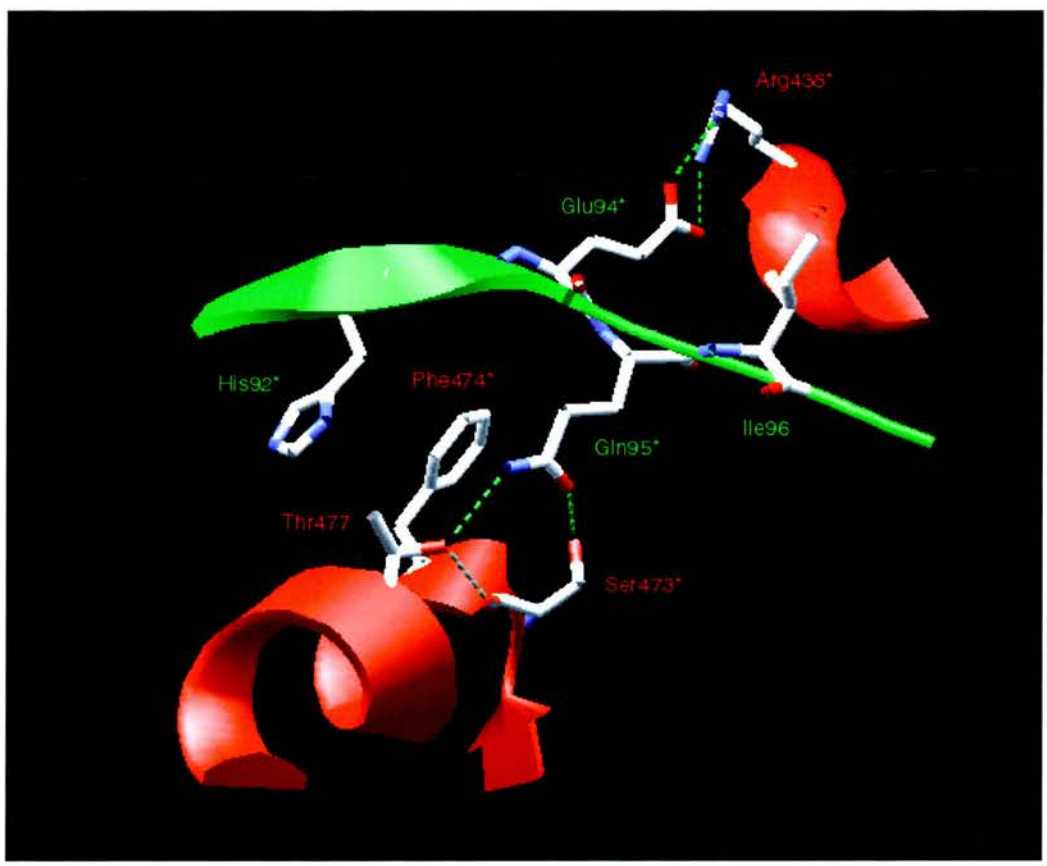
Figure 17. Alignments of the residues proximal to the diglycine motif from a number of Ub-like proteins, and the C-terminal mutants of SUMO-1. Alignments of the C-terminal regions of *S. cerevisiae* Smt3 (ScSmt3), human SUMO-1 (hSUMO-1) and the mutant SUMO-1 proteins (altered residues shown in blue), *S. pombe* Smt3 (SpSmt3), and the human sequences for Ub (hUb) and NEDD8. Homologous and identical residues are shown as capitals, and capital bold respectively, and residues known to make contacts with Ulp1 through side-chain interactions, indicated by black dots. Sequences cleaved by the SUMO-1 and/or Smt3 proteases are shown divided by the red line, lower efficiencies are shown by a broken line, and those known not to be cleaved with no line.

which Phe474 of Ulp1 sits, forming a large VDW contact (Mossessova and Lima, 2000). The importance of the VDW interactions are demonstrated by the mutation of Phe474 to Ala, which is lethal to yeast (Mossessova and Lima, 2000). Furthermore, as both Ser473 and Phe474 are functionally conserved in SENP1 (Gong *et al.*, 2000b) and other Smt3 proteases (Mossessova and Lima, 2000) (see Figure 18), it is probable that similar interactions are present in SUMO-1/hydrolase complexes, which were disrupted by the mutation of Gln94 to Leu.

Interestingly, although no C-terminal hydrolase data could be obtained for E93R-SUMO-$1_{(1\text{-}101)}$, the double mutation containing both E93R and T95R mutations is over ten times more resistant to hydrolysis than WT-SUMO-1 or the T95R mutant alone. This suggests that the E93R mutation either plays the more significant role of the two residues in the inhibition mechanism, or that the two mutations together act synergistically to partially stabilize the protein. The Smt3 equivalent to Glu93, Glu94 forms a salt-bridge with Arg438 from Ulp1 in the complex (Mossessova and Lima, 2000) (see Figure 18). Arg438 appears to be functionally conserved as Lys in some SENP proteases (Gong *et al.*, 2000b), suggesting that the salt-bridge may also be common to the SENPs. Concurrent with the *in vivo* data, the triple SUMO-1 mutant, RLR-SUMO-1 was the least susceptible to hydrolysis, with an initial rate twenty times slower than the double mutant, and over 80 times slower than the WT. This represents a simple multiple inhibition provided by the twenty times and four times slower rates of RR-SUMO-1 and Q94L-SUMO-1 put together.

Figure 18. Ribbon diagrams showing the Smt3-Ulp1 complex and the specific hydrogen bonds made between the C-terminal side-chains of Smt3 and residues on two helices of Ulp1. The upper panel shows a ribbon diagram of the crystal structure of Smt3 (yellow and green) in complex with an active fragment (403-621) of the Smt3-specific protease Ulp1 (magenta and red). Red ribbons correspond to those in the lower panel which show a zoom view of the regions known to contain residues that make specific hydrogen bonds or Van der Waals contacts with side-chains of residues in the Smt3 C-terminus (green). Involved residues from Smt3 (green) and Ulp1 (red) are labelled where * denotes residues either identical or homologous to those in SUMO-1 and the SUMO-1 protease SENP1. Figures generated using Swiss PDB viewer v3.6 (Guex *et al.*, 1999).


C
N
N


Arg438*
Glu94*
His92*
Phe474*
Ile96
Gln95*
Thr477
Ser473*

Interestingly, the results of the analysis of the SUMO-1 C-terminal mutants in the *in vitro* isopeptidase assay were not wholly in agreement with the C-terminal hydrolase data. Although, only representing a single time point in the deconjugation process (1h), only the triple mutant of SUMO-1 was significantly less susceptible to removal from PML by isopeptidase activity/activities in the HeLa extract. The RLR-SUMO-1-PML conjugates were significantly more stable than those of the WT, with only approximately 32% of the conjugated RLR-SUMO-1 removed from PML in the same period of time 75% of WT-SUMO-1-PML conjugates were removed. Although the protease content of the HeLa fraction is unknown, and it is not yet clear whether a single SENP is responsible for the maturation of newly synthesised SUMO-$1_{(1\text{-}101)}$, or that individual SENP proteins have distinct or overlapping substrate specificities, in general these results show that the isopeptidase activity/activities present in the HeLa cell extract is not as strongly influenced by the mutation of residues 93-95 of SUMO-1 as is the C-terminal hydrolase activity. One explanation for this observation is that when removing SUMO-1 from conjugated substrates, the SUMO-1-isopeptidase complex is stabilised by interactions with the conjugated protein itself, and thus the interactions between the protease and SUMO-1 itself would be relatively less important. The unexpected result that the mutation of Thr95 to Arg which has no effect on C-terminal hydrolase activity, appears to marginally reduce its ability to be removed from PML, may also be explained in terms of PML interactions. Although as discussed earlier it is unlikely that this mutation

directly affects protease interactions, it is possible that when conjugated to PML it may interfere indirectly with PML-protease contacts, thus slightly destabilizing the complex.

Overall, in comparison with the conjugation data (Sections 3.2), these results imply that the residues proximal to the diglycine motif in SUMO-1 are not as important for efficient proteolytic processes, as they are conjugation. This confirms that Ubc9 and SUMO-1 proteases rely, in part, on differing contacts with SUMO-1 when catalysing their respective reactions.

3.4. A variant loop within Ubc9 core domain may act as a recognition element for the SUMO-1 substrate consensus motif

3.4.1. Results

3.4.1.1. A Ubc9 variant loop is inhibitory to SUMO-1-thioester formation

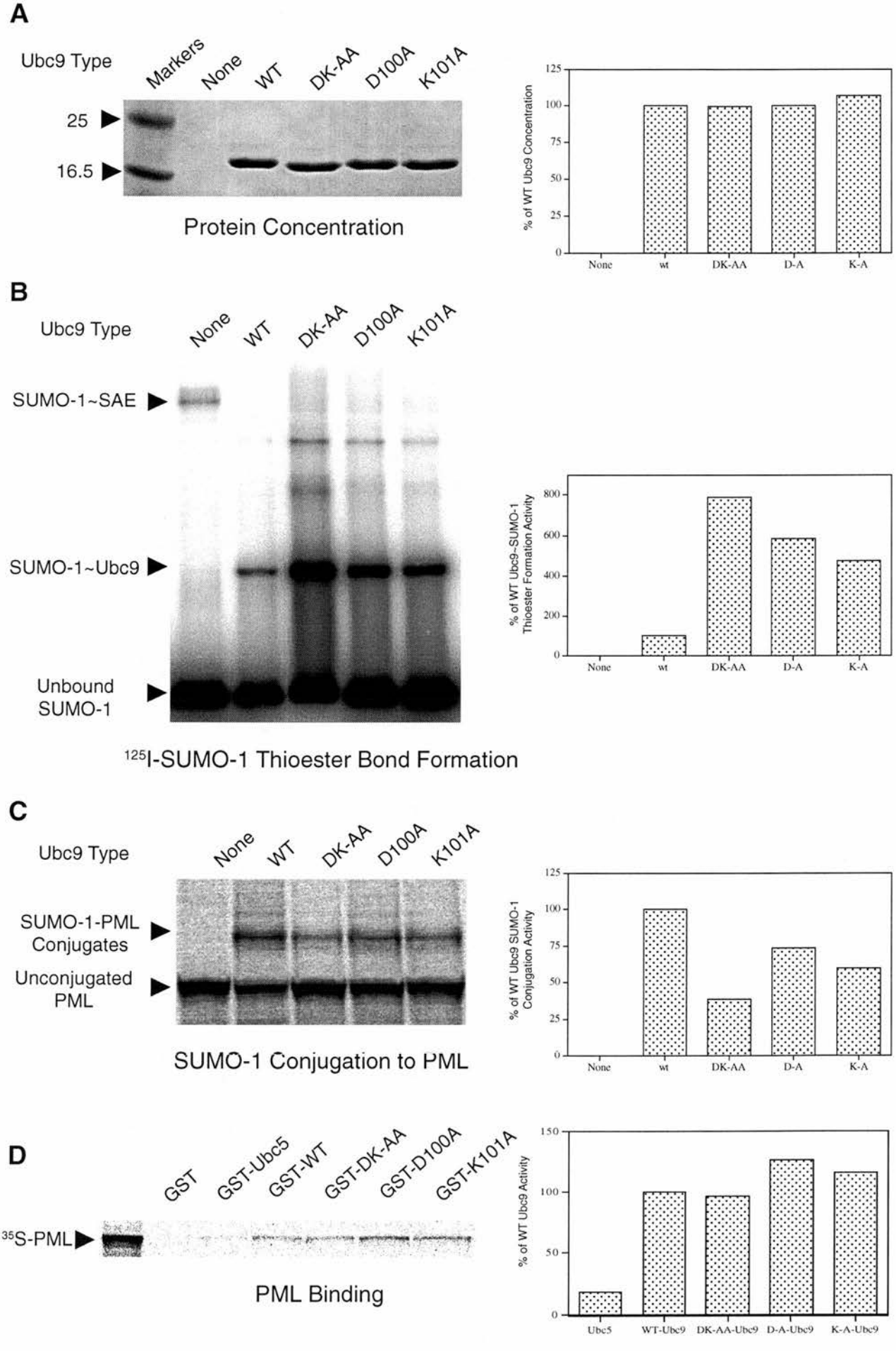
The E2 conjugating enzymes of both Ub and UBL-conjugation pathways have a core domain of approximately 150 amino-acids that show at least 25% sequence identity. All E2 enzymes can be divided into four classes depending on whether they contain extra C-terminal or N-terminal extensions, neither, or both (Jentsch, 1992). These extensions have been proposed to confer specificity for recognition of substrates or E3s (Silver *et al.*, 1992), or to provide a localisation signal (Yang *et al.*, 1997). The SUMO-1-conjugating enzyme Ubc9 has neither a C-terminal nor N-terminal extension, yet without the aid of E3 enzymes, catalyses the transfer of SUMO-1 to lysine residues within a specific subset of protein substrates which almost always present the ΨKXE consensus motif (where ψ is a hydrophobic amino-acid). Sequence alignments with other E2 enzymes (Figure 4) reveals that some conjugating enzymes contain an insertion proximal to the catalytic Cys residue. Ubc9 contains such an insertion of only two amino-acids, Asp100 and Lys101. To functionally characterise this short loop, firstly three site-directed mutants of Ubc9, two single mutations, D100A-Ubc9, K101A Ubc9, and one double DK-AA-Ubc9 were cloned using PCR. DNAs were cloned into pGEX-2T plasmids, and

proteins expressed in *E. coli* as GST-fusion proteins. To functionally characterise these variants the Ubc9 proteins cleaved from GST were firstly tested for their ability to accept ^{125}I-SUMO-1-thioesters from the SUMO-1 E1 activating enzyme (SAE1/2) in the previously described ^{125}I-SUMO-1 thioester assay (Desterro *et al.*, 1999). Equimolar quantities of the Ubc9 (Figure 19A) proteins clearly show differing abilities to form ^{125}I-SUMO-1 thioester bonds (Figure 19B). In this 'end-point' assay where the thioester reactions were incubated for 30 min, the double mutant DK-AA-Ubc9 appears to be hyper-active, and forms almost eight times the quantity of ^{125}I-SUMO-1 thioester bonds than the WT protein. The enhanced activities of the single mutants were not quite as dramatic as that for the double, with D100A-Ubc9 and K101A-Ubc9 forming over five-times and four-times respectively the amounts of ^{125}I-SUMO-1-thioesters generated by WT-Ubc9. This suggests that under these *in vitro* conditions the side chains from the D100-K101 loop in Ubc9 are actually inhibitory to SUMO-1 thioester bond formation.

3.4.1.2. Loop mutants of Ubc9 have impaired conjugation activity, but still bind substrate

To identify the overall conjugation activity of the Ubc9 mutants the *in vitro* SUMO-1 conjugation assay was employed. Again equimolar quantities of Ubc9 proteins were compared, this time for their ability to conjugate SUMO-1 to *in vitro* translated ^{35}S-PML (Figure 19C). In contrast to the thioester experiment, the loop mutations of Ubc9 appear to inhibit

Figure 19. Functional comparison between WT-Ubc9 and mutants of the variant loop proximal to the catalytic Cys93. Mutants of Ubc9 of D100A K101A (DK-AA), D100A and K101A were generated using PCR and cloned into the GST-fusion expression vector pGEX-2T. Expressed proteins were purified and thrombin cleaved before dialysis and protein concentration assay. **A.** 2μg of each of the mutant proteins and the WT-Ubc9 were fractionated by 12% Polyacrylamide gel electrophoresis before Coomassie staining and destaining and gel photography. Molecular weight markers are indicated. **B.** Equal quantities of each Ubc9 mutant was assayed for the ability to form thioester bonds with SUMO-1 using the previously described SUMO-1 thioester assay (Desterro *et al.*, 1997). Assays were incubated for 30 mins before addition of 1M urea disruption buffer and fractionation by 15% polyacrylamide gel electrophoresis, drying and exposure to phosphorimage screening. **C.** The overall ability for each mutant to conjugate SUMO-1 to the target ^{35}S-PML was assayed using the previously described SUMO-1 conjugation assay (Desterro *et al.*, 1998). Reactions were again analysed by polyacrylamide gel electrophoresis, drying and phosphorimaging. **D.** Recombinant GST-fusion Ubc9 proteins GST-Ubc5 and GST alone were bound to glutathione-sepharose beads. Standardised quantities of each bead sample were incubated with ^{35}S-PML and washed as detailed in the Materials and Methods section. ^{35}S substrate associated with the proteins on the beads were released by addition of SDS β-mercaptoethanol sample buffer before boiling for 1 min and fractionation by 10% polyacrylamide gel electrophoresis, drying and exposure to phosphorimage screening. Relative band intensities for each experiment were calculated using Multianalyst (MultiAnalyst V 1.02) or MacBAs (MacBas V 2.5) quantification programs and represented graphically (right).


A
Ubc9 Type
Markers
None
WT
DK-AA
D100A
K101A
25
16.5
Protein Concentration
% of WT Ubc9 Concentration
125
100
75
50
25
0
None
wt
DK-AA
D-A
K-A
B
Ubc9 Type
None
WT
DK-AA
D100A
K101A
SUMO-1~SAE
SUMO-1~Ubc9
Unbound SUMO-1
125I-SUMO-1 Thioester Bond Formation
% of WT Ubc9~SUMO-1 Thioester Formation Activity
800
600
400
200
0
None
wt
DK-AA
D-A
K-A
C
Ubc9 Type
None
WT
DK-AA
D100A
K101A
SUMO-1-PML Conjugates
Unconjugated PML
SUMO-1 Conjugation to PML
% of WT Ubc9 SUMO-1 Conjugation Activity
125
100
75
50
25
0
None
wt
DK-AA
D-A
K-A
D
GST
GST-Ubc5
GST-WT
GST-DK-AA
GST-D100A
GST-K101A
35S-PML
PML Binding
% of WT Ubc9 Activity
150
100
50
0
Ubc5
WT-Ubc9
DK-AA-Ubc9
D-A-Ubc9
K-A-Ubc9

the overall conjugation to PML. In fact, the DK-AA double mutant has the lowest efficiency, accumulating only ~35% of the SUMO-1-PML conjugates formed by the WT-Ubc9. The individual mutations are less affected by their amino-acid substitutions, as D100A-Ubc9 forms ~72%, and K101A-Ubc9 ~59% of the total WT-Ubc9 catalysed SUMO-1-PML conjugates. To determine whether or not loop mutations alter the ability of Ubc9 to bind protein substrates, a GST-pulldown assay was used. GST-fusion proteins of WT-Ubc9, the mutants and the Ub conjugating enzyme Ubc5 were bound to glutathione-sepharose beads, and analysed for protein mass:bead volume ratios. By using different bead volumes, equimolar quantities of protein were used to bind ^{35}S-PML (Figure 19D). Surprisingly, GST-DK-AA appeared to bind ^{35}S-PML with the same affinity as GST-WT-Ubc9, while GST did not pulldown PML at all. Furthermore, the D100A and K101A-Ubc9 variants appeared to bind PML to a marginally higher degree than either WT or DK-AA proteins, suggesting that the loop is not majorly involved in the global substrate recognition mechanism. The same results were also seen with similar experiments using ^{35}S-p53 (data not shown).

3.4.2. Discussion

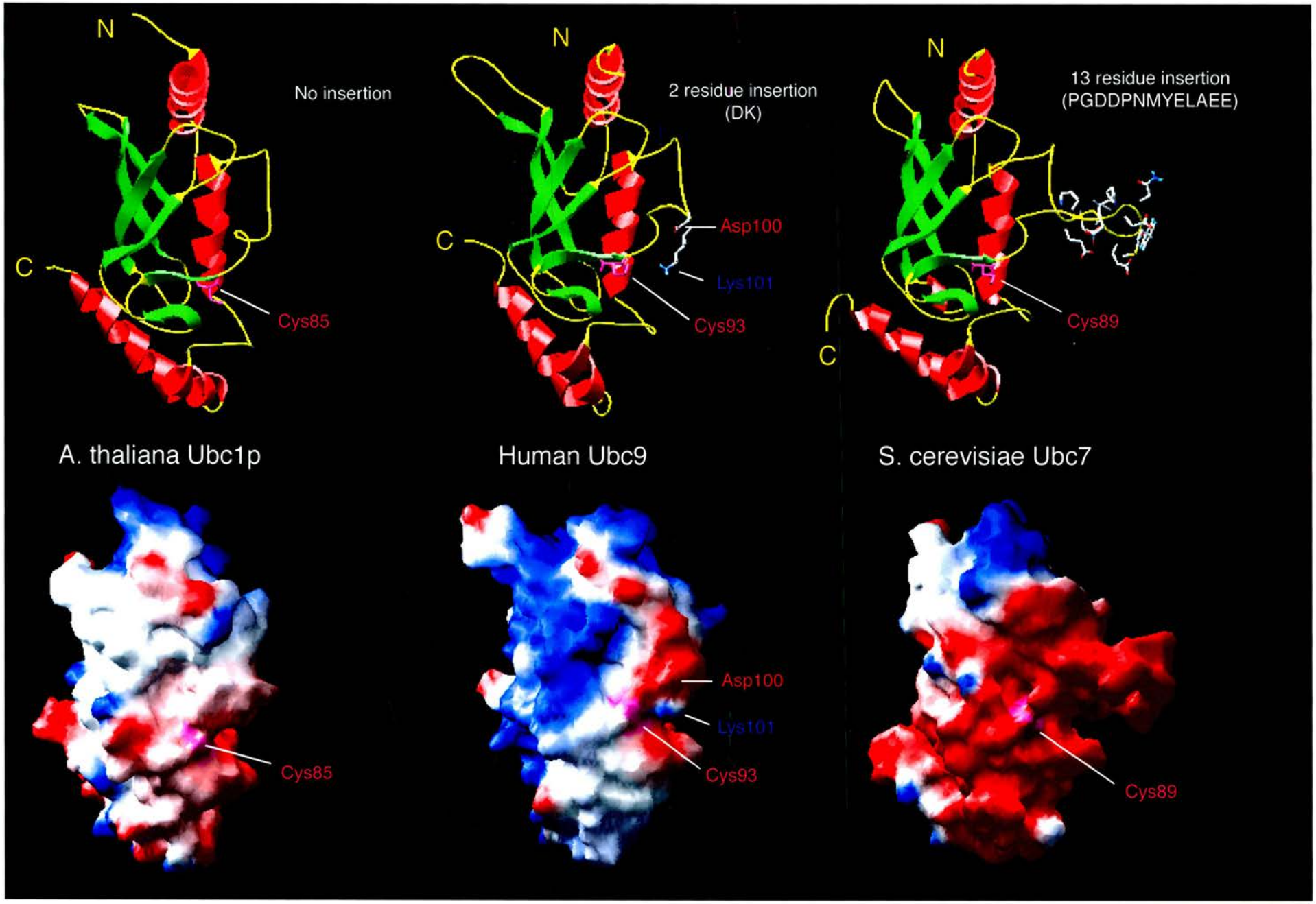
Mutational studies of E2 enzymes have in the past generally focused on the catalytic Cys residue, or C- and N-terminal extension from the ~150 residue conserved core domain. However, sequence alignments of E2 enzymes reveal that some conjugating enzymes contain small insertions proximal to the active-site Cys residue (Figure 20), which have conversely, been relatively poorly investigated. Importantly, E2 enzymes with variant insertions tend to have functional differences from other closely related E2 family members. For example, the yeast Ubc7 enzyme which, like its human homologue, contains a 13 amino-acid insertion 6 codons upstream from the Cys residue, is unable to support E3-dependent conjugation to endogenous proteins, unlike other similar members of the class I E2s (Lin and Wing, 1999). A recent report partially characterising this loop in the rabbit homologue of Ubc7 known as $E2_{17K}$, indicated that it negatively influences binding with an E1-Ub thioester complex, and that although the core domain of the protein plays an important role in defining substrate specificity, the loop plays a subtle role in substrate selectivity (Lin and Wing, 1999). Structural comparisons between the plant Ubc1 and yeast Ubc7 (Figure 21 left and right, upper images) show that the insertion manifests itself as a loop that protrudes from the core E2 domain spatially proximal to the active site Cys residue. However in contrast, the side-chains of Asp100 and Lys101 from the two residue insertion in Ubc9 (Figure 20), actually point in towards the active-site cysteine (Figure 21

Figure 20. Sequence alignment of E2 conjugating enzymes from both humans, rabbit and yeast showing the various insertions found in different enzymes proximal to the catalytic Cys residue. Alignment shows the human Ub E2 UbcH7 (accession number NP_003334), the plant *Arabidopsis thaliana* Ub E2, and both *S. cerevisiae* (Sc) and human (h) homologues of the NEDD8/Rub1 conjugating enzyme Ubc12 (accession numbers NP_13409 and XP_009365), the SUMO-1 conjugating enzyme Ubc9 (accession numbers NP_010219 and XP_007786) and the 13 residue loop-containing Ub E2s Ubc7 (accession numbers NP_003334 and 2981900) and the rabbit homologue rUbc7 (accession number AAC69605). The catalytic Cys residues are shown in red, homologous residues in capitals, identical residues in bold capitals, and loop insertion residues in blue.

central, upper image). Furthermore, electrostatic surface potential comparisons, show that while the negative loop from Ubc7 generates a negatively charged extension pointing out from the catalytic Cys residue, Lys101 actually presents a localised positive charge proximal to Cys93, in a region of general negativity (Figure 21 lower images).

Consistent with the Ubc7 studies, the mutation of residues D100-K101 in the variant loop in human Ubc9 results in mutants augmented in their ability to form SUMO-1 thioester bonds. This shows the loop to be inhibitory to maximal thioester formation efficiency, and thus, is likely to either stearically hinder the transfer of SUMO-1 from the SAE2 subunit of the heterodimeric SAE1/SAE2 SUMO-1-activating enzyme, or it impedes the physical interaction between E1 and E2. Furthermore, although Ubc9 loop mutants had increased thioester formation efficiency, their ability to then conjugate SUMO-1 to PML *in vitro* was shown to be severely compromised. In fact, in relative terms of thioester activity, the double mutant DK-AA-Ubc9 only retains approximately 4% of the WT efficiency in catalysing the final isopeptide bond with target lysine residues. The thioester results show that this poor substrate transfer is not an artefact of an unusually high affinity of the mutants for the E1-SUMO-1 thioester complex, as E1 and E2 dissociate after transesterification of the SUMO-1 molecule (Figure 19B). Thus, as pulldown experiments revealed no significant difference in overall substrate affinity, it can be proposed that like the 13 residue insertion in Ubc7, the Ubc9 loop may in fact play a subtle role in substrate recognition.

Figure 21. Ribbon and spacefill diagrams of plant and yeast Ub conjugating enzymes Ubc1p and Ubc7, and the human SUMO-1 conjugating enzyme Ubc9. Ribbon diagrams show helix structures in red, sheet structures in green, random coils in yellow, and the indicated active site Cys residues are shown in stick form coloured magenta. All three E2 enzymes present the conserved E2-fold, although Ubc9 and Ubc7 have two, and 13 residue insertions respectively, which are shown with their side-chain structures (Asp100 and Lys101 indicated for Ubc9). Spacefill representations show positively charged and negatively charged regions as blue and red respectively, with active site Cys residues indicated in magenta. All figures created using Swiss PDB viewer v3.6b (Guex *et al.*, 1999). Electrostatic calculations are based upon a dielectric constants of 80.0 (solvent) and 4.00 (protein) and physiological pH.


N
No insertion
C
Cys85
A. thaliana Ubc1p
N
2 residue insertion
(DK)
Asp100
Lys101
Cys93
C
Human Ubc9
N
13 residue insertion
(PGDDPNMYELAEE)
Cys89
C
S. cerevisiae Ubc7
Cys85
Asp100
Lys101
Cys93
Cys89

Interestingly, the Ub-specific E2 Ubc5, although not identified as targeting PML for ubiquitination, appears to weakly bind PML and p53 in the same experiments (Figure 19D). Thus the E2 core domain common to all E2 enzymes independent of modifier specificity, may inherently retain low levels of substrate recognition, although other more specific surface determinants likely define stronger interactions. Bearing this in mind, and considering the size and proximity to Cys93 of the insertion loop in Ubc9, it is possible that the Asp-Lys residues are involved in the direct recognition of amino-acid side-chains proximal to the target lysine and common to all SUMO-1 substrates, namely the ΨKXE SUMO-1 consensus motif. If this is the case, then clearly such an interaction is not essential for the simple docking of E2 to substrate, but it may be required to locate or orient the target lysine with the SUMO-1-Cys93 thioester allowing the efficient formation of the isopeptide bond.

A common feature of Ub E3 enzymes is the small metal-binding motif known as the RING finger domain (Tyers and Willems, 1999). The recently reported crystal structure of a human Ub-conjugating enzyme termed UbcH7 (although not functionally related to yeast or rabbit Ubc7 and has no variant loop insert) in complex with its RING family E3, c-Cbl and a small fragment of substrate peptide, has revealed the important contacts made between E2 and E3 enzymes (Zheng *et al.*, 2000). Specifically, UbcH7 contacts Cbl via the RING domain inserting two coil structures close to the N-terminus of UbcH7 into a groove on the surface of Cbl. One of these coils corresponds to the region into which the variant insertions

are made in Ubc9 and Ubc7. Interestingly, it has been documented that a number of proteins known to interact with Ubc9 do so via specific secondary structures, such as the helix-loop-helix motifs of TEL (Chakrabarti *et al.*, 2000) and MITF proteins (Xu *et al.*, 2000). Thus it is tempting to suggest that regions equivalent in Ubc9, to those known to be involved in E3 binding by UbcH7, may specifically interact with grooves provided by secondary structures inherent in SUMO-1 substrates analogous to Ub-E2-E3 binding, thus explaining the apparent lack of requirement for E3 enzymes in the SUMO-1 conjugation pathway. The absence of the sequence containing the target Lys in the Cbl-UbcH7 structure, has meant that we still do not know the exact contacts made between E2 and substrate (and modifier) in the catalytic complex. The structural elucidation of a complex containing a conjugating enzyme with a fragment of interacting template substrate would considerably improve our knowledge of the poorly understood mechanisms of the final step of the conjugation of Ub and UBLs to specific proteins. Due to the apparent absence of requirement for E3 activities by the SUMO-1 conjugation pathway, a Ubc9/SUMO-1/substrate catalytic complex would be an ideal candidate.

3.5. SUMO-2 and SUMO-3 are Conjugated to Protein Substrates by SAE1/2 and Ubc9

3.5.1. Results

3.5.1.1. SUMO-2 and SUMO-3 modify a different spectrum of proteins from that modified by SUMO-1

Although the sequences of SUMO-2 and SUMO-3 share approximately 50% identity with SUMO-1 (Figure 3, Table 3), functional differences between SUMO-1 and SUMO-2/-3 have been reported (Saitoh and Hinchey, 2000). To compare the substrates modified by SUMO-1, SUMO-2 and SUMO-3 *in vivo*, HA epitope tagged versions of these proteins were expressed transiently in COS7 cells, and the subsequent spectrum of modified proteins analyzed by Western-blotting with an anti-HA specific monoclonal antibody. While the pattern of modified proteins is broadly similar between SUMO-2 and SUMO-3. They both display a significantly different profile from that observed with SUMO-1 (Figure 22). In particular six immunoreactive species in the 85-200 kDa range differ between HA-SUMO-1 and HA-SUMO-2 or HA-SUMO-3 transfected cell lysates. Conjugated species labeled 2 and 4 are present in cells expressing HA-SUMO-1, but are undetectable in cells expressing HA-SUMO-2 and HA-SUMO-3. Labelled species 6 is present in cells expressing HA-SUMO-2 and HA-SUMO-3, but is almost undetectable in cells expressing HA-SUMO-1.

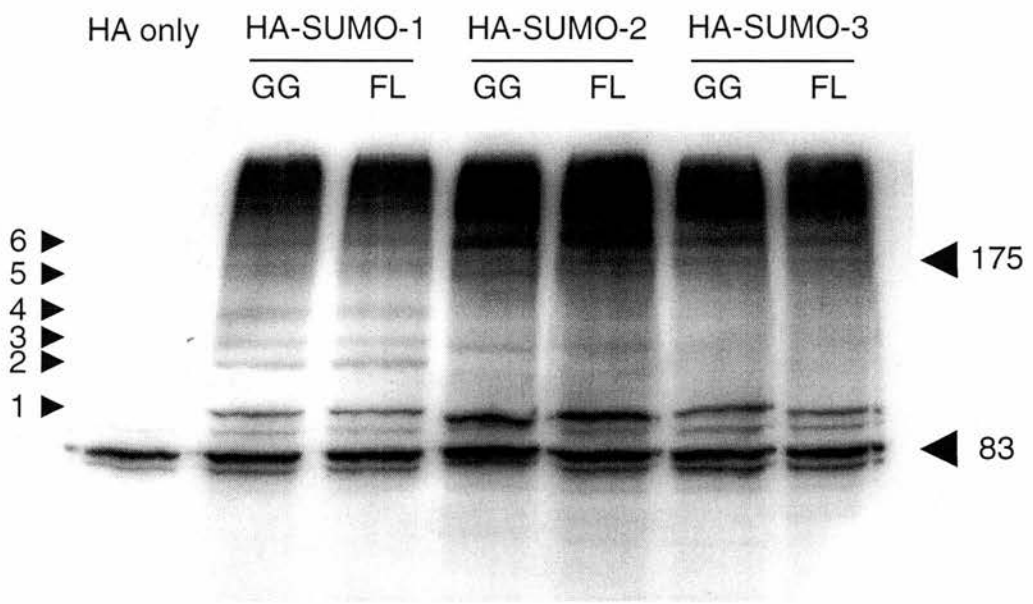


Figure 22. The spectrum of proteins conjugated by SUMO-1 in COS7 cells differs from those conjugated by SUMO-2 and SUMO-3. COS7 cells were transiently transfected with pCDNA3 vectors containing HA-Tag only (HA only), HA-SUMO-1, HA-SUMO-2 or HA-SUMO-3 full length (FL) or truncated (GG). Cells were incubated for 36h post-transfection before lysis and fractionation by 8% polyacrylamide gel electrophoresis. HA reactive species were detected by Western-blotting. Clearly visible SUMO-protein conjugates are numbered 1-6, and molecular weight markers are indicated.

It is also clear that the pattern of SUMO conjugates is the same in cells expressing either the full-length (FL) or shorter (GG) form of each SUMO protein. This indicates that the over-expressed FL proteins are being efficiently processed by the specific protease (or proteases). This is confirmed by analysis of the low molecular weight region of blots from 15% polyacrylamide gels, which show efficient cleavage of these precursor constructs to their active forms (data not shown).

3.5.1.2. SUMO-1, SUMO-2 and SUMO-3 each form SAE1/2-dependent thioester bonds with Ubc9 in vitro

Although SUMO-1 and SUMO-2/-3 share significant sequence identity (Figure 23A), they appear to modify different subsets of proteins *in vivo* (Figure 22). It is therefore important to determine if SUMO-1 and SUMO-2/-3 are linked to substrates by the same conjugation system. The ability of SUMO-1, SUMO-2 and SUMO-3 to be utilised by SAE1/SAE2 and Ubc9 was therefore tested in a previously described *in vitro* SUMO-1 thioester assay (Desterro *et al.*, 1997), modified to facilitate analysis by Western-blotting. This was necessary as SUMO-2 and SUMO-3 could not be labelled with ^{125}Iodine due to the lack of exposed tyrosine residues. SUMO-1, SUMO-2 and SUMO-3 were expressed and purified as both the full-length precursors and the shorter active forms (exposing the C-terminal diglycine motif). GST-fusion and thrombin-cleaved versions of each protein were analysed by SDS PAGE and Coomassie Blue staining (Figure 23B). The thrombin cleaved proteins were essentially

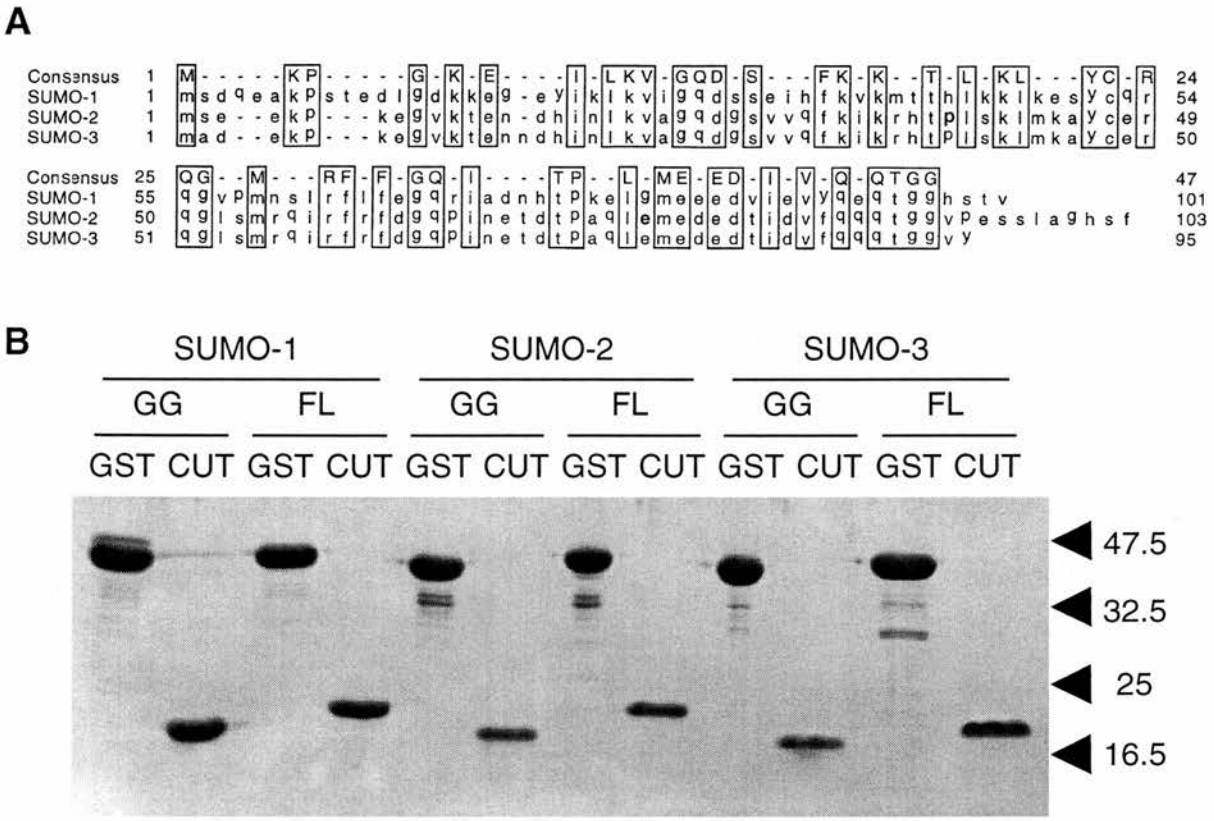


Figure 23. Primary sequence alignment of SUMO-1, SUMO-2 and SUMO-3, and a Coomassie stained gel of recombinant proteins. A. Primary sequence alignment of SUMO-1, SUMO-2 and SUMO-3, showing the consensus sequence. Boxed residues are identical. Two codon changes were identified between the published sequence for SUMO-2/SMT3A (accession number X99584) and the sequence obtained from our clones. These code for amino-acid changes S38P and R76E, which are shown in bold type, and are conserved in SUMO-3. B. Coomassie stained gel of SUMO-1/-2/-3 protein constructs. GST-fusions were eluted with glutathione, while unfused SUMO proteins were released from the beads by thrombin cleavage. 5µg of each GST-fusion (GST) and thrombin cleaved (CUT) samples of both the truncated (GG) and full-length (FL) constructs of SUMO-1, SUMO-2 and SUMO-3 were fractionated by 15% polyacrylamide gel electrophoresis, before staining with Coomassie Brilliant blue R250. Molecular weight markers are indicated. Differences in apparent band intensities can be explained by the varying ability of each SUMO protein to react with Coomassie R250.

homogeneous, and their mass determined by mass spectrometry, corresponded to that predicted from the cloned sequences. Recombinant SUMO-1, SUMO-2 or SUMO-3 was incubated in the assay with Ubc9, IPP and ATP in the presence or absence of recombinant SAE1/SAE2. A fraction of the terminated reactions were separated by electrophoresis in 15% polyacrylamide gels containing SDS, before Western-blot analysis. Antibodies used were specific for Ubc9, SUMO-1 or SUMO-2/-3 (Figure 24). Analysis of the SUMO-1 assays (upper panels) indicates that in the absence of SUMO-1 (None) or in the presence of the inactive full-length protein (FL), no SAE1/SAE2 dependent Ubc9 or SUMO-1 antibody reactive species are generated. The most prominent species in the Ubc9 blot (right) at approximately 35 kDa is likely to be Ubc9~Ubc9 disulfide linked dimers, which are present in every reaction, and are reduced in the presence of β-mercaptoethanol (data not shown). If the GG form of SUMO-1 is included an SAE1/SAE2-dependent species is generated below the Ubc9~Ubc9 dimers, at the apparent molecular weight of 32kDa, which is also reactive to the SUMO-1 antibody. In the presence of the GST-fusion construct of the GG protein (GST-GG), an SAE1/SAE2-dependent species is generated at approximately 55kDa, which is also reactive to both Ubc9 and SUMO-1 antibodies. Again, both these species are sensitive to reducing agent (data not shown). These results are consistent with the formation of SUMO-1~Ubc9 and GST-SUMO-1~Ubc9 thioester complexes via the C-terminal glycine of SUMO-1 and Ubc9 as described previously (Desterro *et al.*, 1997). The same observations can be made with respect to SUMO-2 (middle panels) and SUMO-3 (lower panels) indicating that SUMO-1, SUMO-2 and

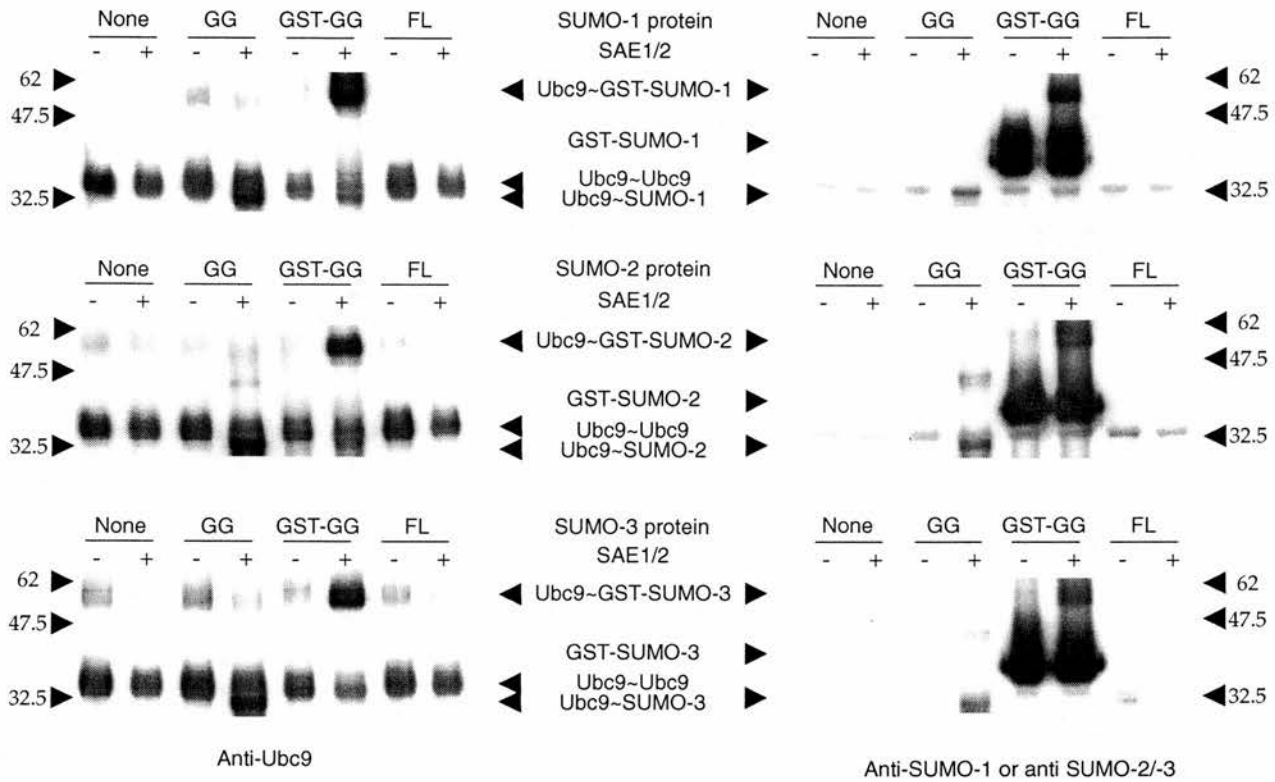


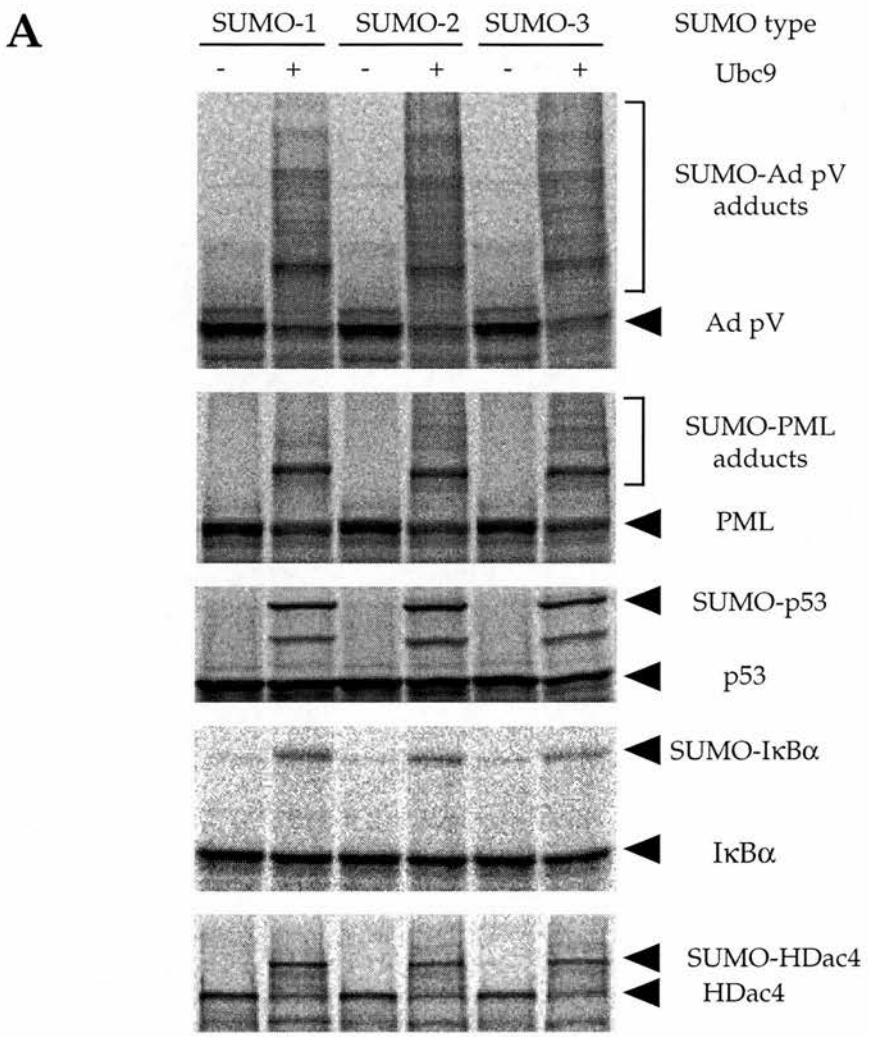
Figure 24. SUMO-1, SUMO-2 and SUMO-3 all form SAE1/2-dependant thioester bonds with Ubc9 *in vitro*. SUMO-1 (top panels), SUMO-2 (middle panels) or SUMO-3 (lower panels) proteins were incubated in the SUMO thioester assay as described in the experimental procedures in the presence (+) or the absence (-) of recombinant SAE1/2. The full length (FL), the truncated (GG) and the GST-fusion GG (GST-GG) constructs of each protein were tested for their ability to form SAE1/2-dependent thioester bonds with Ubc9. A fraction of each 10ml reaction mixture was fractionated by 15% polyacrylamide gel electrophoresis. Gels were analysed by Western-blotting using antibodies specific to Ubc9 (left panels), SUMO-1 or SUMO-2/-3 (right panels). The thioester species formed between the GG and GST-GG versions of SUMO-1, SUMO-2 and SUMO-3 are indicated, as are Ubc9 disulfide dimers and the unbound GST-GG proteins.

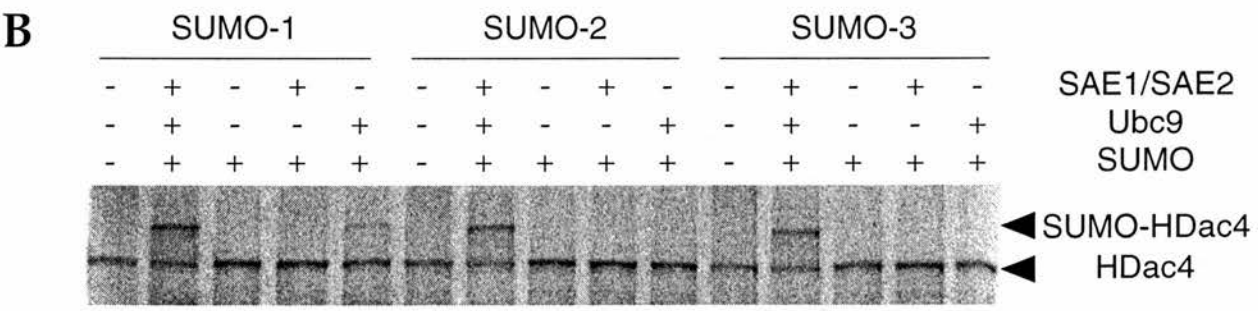
SUMO-3 form SAE1/SAE2 dependent thioester bonds with Ubc9. As all of the proteins used are purified, recombinant proteins from bacteria, this rules out the requirement for other cellular proteins in the formation of Ubc9 thioesters with the three SUMO species.

3.5.1.3. Conjugation of SUMO-1, SUMO-2 and SUMO-3 to the same substrates in vitro

To date a number of substrates have been identified as targets for the SUMO-1 conjugation pathway, although only PML and RanGAP1 have been shown to be conjugated to SUMO-2/-3 (Kamitani *et al.*, 1998a; Saitoh and Hinchey, 2000). To identify any difference in substrate specificity, the previously described SUMO conjugation assay was employed (Desterro *et al.*, 1997). The recombinant GG SUMO proteins were assayed for Ubc9-dependent conjugation to a number of known SUMO-1 substrates labelled with ^{35}S-methionine. These included substrates known to be modified on a single lysine (IκBα, p53 and HDac 4), and substrates that are modified at multiple lysines (PML and adenovirus type 2 protein V). The presence of higher molecular weight ^{35}S-labelled species for all three SUMO proteins in the presence of Ubc9 indicates that SUMO-2 and SUMO-3 are as efficiently conjugated as SUMO-1 to all substrates tested (Figure 25A). In the case of HDac 4, conjugation with all SUMO species tested was dependent on the presence of SAE1/SAE2 and Ubc9 (Figure 25B)

Figure 25. SUMO-1, SUMO-2 and SUMO-3 are conjugated *in vitro* to the same substrates in an SAE1/SAE2, Ubc9-dependent reaction. (A) ^{35}S-methionine labelled IκBα, adenovirus type 2 protein V (Ad pV), promyelocytic leukaemia protein (PML), Histone Deacetylase (HDac 4), and p53 were individually incubated in the assay for SUMO conjugation as described in the Materials and Methods, before fractionation by 8% polyacrylamide gel electrophoresis, staining, destaining and drying. (B) SUMO-1, SUMO-2 and SUMO-3 conjugation to Hdac 4 was tested for its dependency on SAE1/SAE2 and Ubc9 as above. Samples were fractionated by 8% polyacrylamide gel electrophoresis before staining, destaining and drying. Gels were analysed by phosphorimaging. Both conjugated and unconjugated forms of the substrates are indicated.

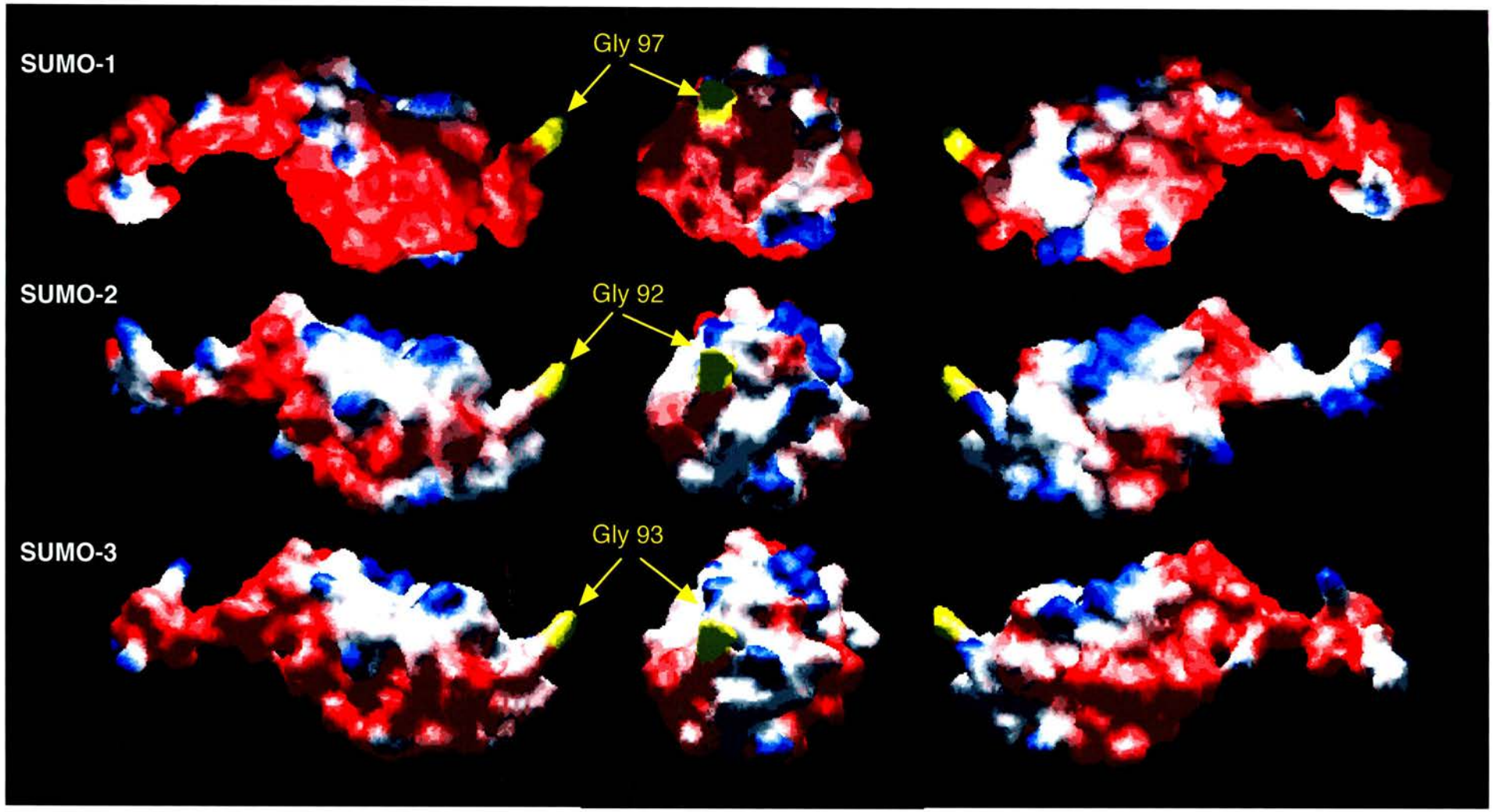

A
SUMO-1
SUMO-2
SUMO-3
SUMO type
- + - + - +
Ubc9
SUMO-Ad pV adducts
Ad pV
SUMO-PML adducts
PML
SUMO-p53
p53
SUMO-IκBα
IκBα
SUMO-HDac4
HDac4


B
SUMO-1
SUMO-2
SUMO-3
- + - + - - + - + - - + - + - SAE1/SAE2
- + - - + - + - - + - + - - + Ubc9
- + + + + - + + + + - + + + + SUMO
SUMO-HDac4
HDac4

3.5.2. Discussion

The results reported here indicate that while SUMO-1 and SUMO-2/-3 are present in conjugates with different proteins *in vivo*, they are conjugated with equal efficiency to the same protein substrates *in vitro*. As with SUMO-1, SUMO-2/-3 conjugation is mediated by the same enzymatic cascade involving SAE1/SAE2 and Ubc9. This is consistent with previous work (Gong *et al.*, 1999) which suggested that SAE1/SAE2 could form thioesters with SUMO-1, SUMO-2 and SUMO-3. As SUMO-1 and SUMO-2/-3 share over 50% sequence identity it was expected that homology models of SUMO-2 and SUMO-3, based on the NMR structure of SUMO-1 (Bayer *et al.*, 1998), predict tertiary structures essentially identical to SUMO-1. However surface charge distribution and side-chain type are markedly different in a number of areas including that proximal to the diglycine (Figure 26). These differences may be responsible for the distinct functional roles SUMO-1 and SUMO-2/-3. Under the control of as yet unknown mechanisms SUMO-1 modifies a wide variety of cellular proteins with diverse consequences. However, in contrast to SUMO-1 conjugation, and much like ubiquitination, SUMO-2/-3 conjugation activity appears to be under the control of general stimuli such as cell stress (Saitoh and Hinchey, 2000). The significance of this is unclear, but it is conceivable that SUMO-2/-3 conjugation may act as a counter-measure to stress induced Ub-mediated protein degradation (Saitoh and Hinchey, 2000).

Figure 26. Spacefill representations of 3D models of SUMO-2 and SUMO-3 based on the structure of SUMO-1. 3D structures for SUMO-2 and SUMO-3 were predicted by submission of primary sequences (as cloned here) with the PDB co-ordinates for the 3D structure of SUMO-1 (Bayer *et al.*, 1998) to the ExPASy proteomics server of the Swiss Institute of Bioinformatics 'Swiss Model' program (Guex and Peitsch, 1997; Peitsch *et al.*, 1996). 3D backbone comparisons reveal the predicted structures to be essentially identical to SUMO-1 (not shown), while electrostatic surface potential space-fill maps show contrasting charge distributions for SUMO-2/-3 and SUMO-1. Images generated using Swiss PDB viewer v3.6b (Guex *et al.*, 1999). Electrostatic calculations are based upon a dielectric constants of 80.0 (solvent) and 4.00 (protein) and physiological pH.


SUMO-1
Gly 97
SUMO-2
Gly 92
SUMO-3
Gly 93

The issue raised by the experiments reported here is that if SAE1/SAE2 and Ubc9 are used to conjugate SUMO-1, SUMO-2 and SUMO-3 to the same protein substrates, then how does the cell produce unique subsets of proteins that are modified with either SUMO-1 or SUMO-2/-3? One possibility is that additional proteins, equivalent to E3 ligases which participate in ubiquitination, impart specificity to the SUMO modification process *in vivo* and discriminate between SUMO-1 and SUMO-2/3. However there is no absolute requirement for such activity *in vitro*. (Desterro *et al.*, 1999; Okuma *et al.*, 1999) and as yet there is no evidence to support the existence of such proteins *in vivo*, although it is possible that stress-induced E3 enzymes could recognise SUMO-2/-3 targets.

An alternative possibility is that discrimination may not be at the level of SUMO attachment, but activation and/or removal. A large family of cysteine proteases responsible for SUMO processing and removal from substrates have been identified (Li and Hochstrasser, 2000; Yeh *et al.*, 2000). It is entirely possible that these proteins may discriminate between SUMO-1, SUMO-2 and SUMO-3 and in the case of SUMO-2/-3, this activity could be functionally regulated by cell stress.

SUMO-1 and SUMO-2/-3 are predominantly localised to the nucleus (Saitoh and Hinchey, 2000), but while they appear to co-localise in some sub-nuclear structures, their overall distribution is distinct. If SUMO-1 and SUMO-2/-3 are targeted to separate nuclear sites then they may selectively modify proteins that are concentrated at these specific sites. All of these hypotheses are testable and further experimentation will shed light on the functional differences between SUMO-1 and SUMO-2/-3.

4. CONCLUSIONS

The reversible post-translational modification of proteins to alter their function, activity, stability, affinity and cellular localisation is an extremely powerful and ubiquitous cellular tool. Modifying groups themselves are highly diverse in chemical composition and structure, and their attachment to and removal from specific proteins is catalysed by specific conjugating and deconjugating enzymes. The ubiquitin family of protein modifiers has enjoyed heavy academic interest, and over the past few years has increased in size as new members have been discovered. In particular, the Small Ubiquitin-like Modifier SUMO-1 has been identified and its conjugation and deconjugation mechanisms largely characterised. SUMO-1 is conjugated to target proteins by the concerted actions of a SUMO-1-specific E1 enzyme known as SAE1/SAE2, and a dedicated E2 called Ubc9. The removal from substrates is catalysed by one of a number of related SUMO-1-specific proteases, which are also responsible for the maturation of the primary SUMO-1 translation product from the 101 residue protein to the active 97 amino-acid form.

Interestingly, comparisons between the SUMO-1 and ubiquitin pathways reveal a number of consistent observations. Both ubiquitin and SUMO-1 and the enzymes of their conjugation to protein substrates carry relatively high levels of sequence or structural similarity in combination with high degrees of specificity for proteins of their respective pathways. Thus, the question can be asked: How is this specificity manifested in the face of such similarity? This work attempted, in part, to answer this question.

The solution structure of SUMO-1 is a good starting point to help explain any functional distinctions between itself and ubiquitin. The NMR structure of SUMO-1 reveals that the protein contains the characteristic ββαββαβ ubiquitin fold, and is of almost identical backbone structure to ubiquitin accepting a flexible 21 residue N-terminal extension. The major difference between the two proteins is the contrasting surface electrostatic charge distribution, with a number of regions of positivity or neutrality in ubiquitin, being strongly negatively charged in SUMO-1. Molecular models based on SUMO-1 of the two "twin brothers", SUMO-2 and SUMO-3 which these data have shown to be conjugated *in vitro* to the same substrates by SAE1/SAE2 and Ubc9, show patterns of surface charge both in common and contrast with SUMO-1. As SUMO-1 and SUMO-2/-3 appear to have distinct functional roles, their overall conjugated protein profile may either be controlled by different E3 enzymes or proteases from those of SUMO-1, or may be a reflection of alternative cellular localisation. Whatever the explanation for these observations, the differing side-chains exposed on the surface of SUMO-2 and SUMO-3 are almost certain to play a role.

Although relatively unsurprising, the structure of SUMO-1 allowed the identification of a number of regions which may be important for both its functional consequences to substrates and its interaction with specific enzymes. One such region is that next to the C-terminal glycine residues of SUMO-1 and ubiquitin which are involved in the thioester bond formation with the respective E1 and E2 enzymes of the distinct pathways. The C-terminus of SUMO-1, Glu93-Gln94-Thr95-Gly96-Gly97 carries a net

negative charge of -1, while the corresponding region of ubiquitin, Arg72-Leu73-Arg74-Gly75-Gly76 carries a net positive charge of +2. Substitution of these residues in SUMO-1 for those from ubiquitin revealed implications of their involvement at all levels of SUMO-1 metabolism, including maturation, conjugation and deconjugation reactions. Specifically, the efficiency of the maturation process removing the four C-terminal His-Ser-Thr-Val residues, appears to rely in part on Glu93 and Gln94, but has no dependency on the side-chain identity of the residue in position 95. Furthermore, only mutations of all three residues or Glu93 and Thr95 together show any significant inhibition of the deconjugation reaction of the removal of SUMO-1 from substrate proteins.

The interpretation of these results was aided by the published structure of the yeast SUMO-1 homologue Smt3 with its protease Ulp1 (Mossessova and Lima, 2000), which showed that the corresponding residues to Glu93 and Gln94 conserved in Smt3 (Glu94 and Gln95), were involved in stabilising side-chain contacts with residues in Ulp1, while the side-chain group of the equivalent non-conserved residue to Thr95 (Ile96) made no contacts.

Characterisation of the SUMO-1 mutants with respect to the conjugation pathway, revealed that the C-terminal residues were more important to the mechanisms and interactions of these enzymes, than to the proteases. Particularly, Glu93 appears to be required for efficient SAE1/SAE2 thioester bond formation, Thr95 is important for Ubc9 interactions, and Gln94 is involved in the transfer of SUMO-1 from Ubc9 to substrate. Interestingly, Asp100 and Lys101 which form a small non-

conserved two residue insertion close to the catalytic Cys93 of Ubc9, are also involved in the role of SUMO-1 transfer to substrate. The substrates of SUMO-1 conjugation have been relatively well investigated, and the consensus motif ΨKXE where Ψ is any large hydrophobic residue, **K** the target lysine and X any residue (Yeh *et al.*, 2000), is found in all substrates. So it could be tentatively proposed that Gln94 of SUMO-1 together with Asp100 and Lys101 from Ubc9 may be involved in the recognition of residues from the consensus motif, and that while this interaction is important for the bond formation between Gly97 of SUMO-1 and the target lysine, it is not required for efficient substrate recognition by the Ubc9-SUMO-1 thioester complex.

Unfortunately no structural or functional studies to date shed further light on the interactions proposed above. As such it was useful to construct a model simulating the interactions made between SUMO-1, Ubc9 and substrate during catalysis. This model uses the known free solution and crystal structures of SUMO-1 and Ubc9 to from an artificial complex with a small fragment of the substrate p53 (383-390) containing the consensus motif. The molecules were arranged by eye in a simulated thioester complex, bearing in mind the functional consequences of the mutations made, and the requirement of the system for the residues of the consensus motif from the target protein. The model energy was then minimised using GROMOS96 43B1 from the Swiss Pdb viewer software (Guex *et al.*, 1999), and the inter-molecule H-bonds calculated.

It should be noted that although the structure shows SUMO-1 only contacting Ubc9 via the C-terminal residues with the body of the protein

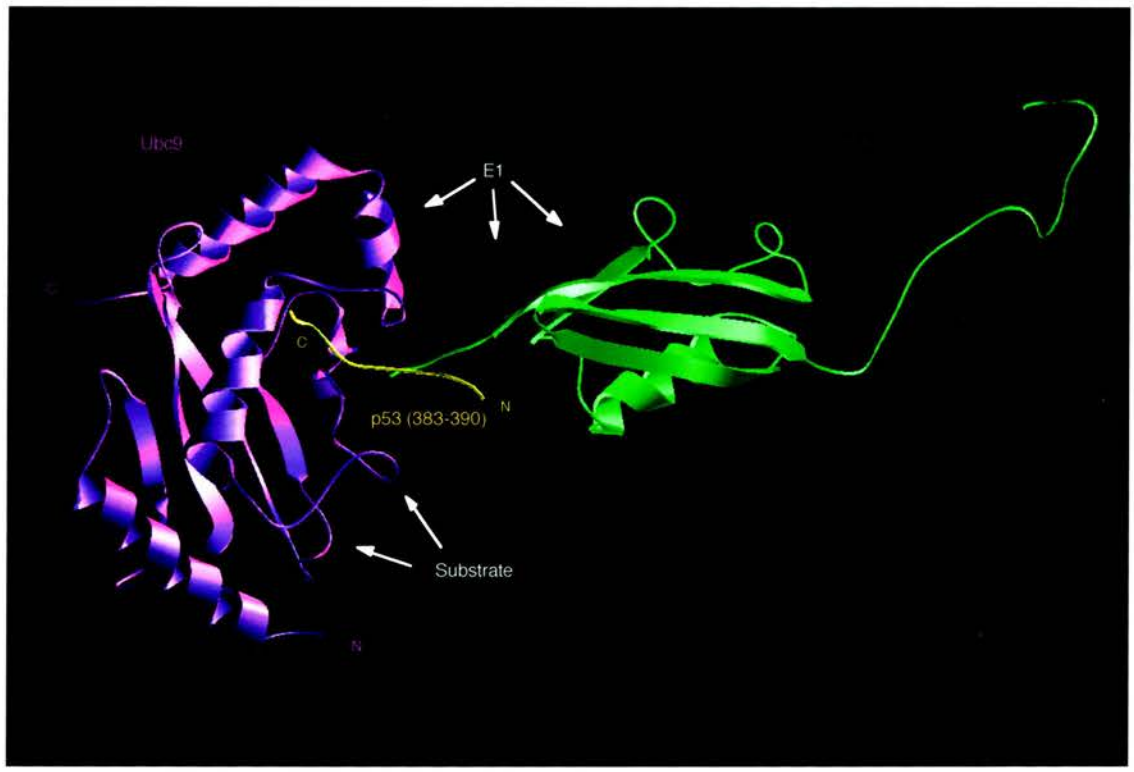
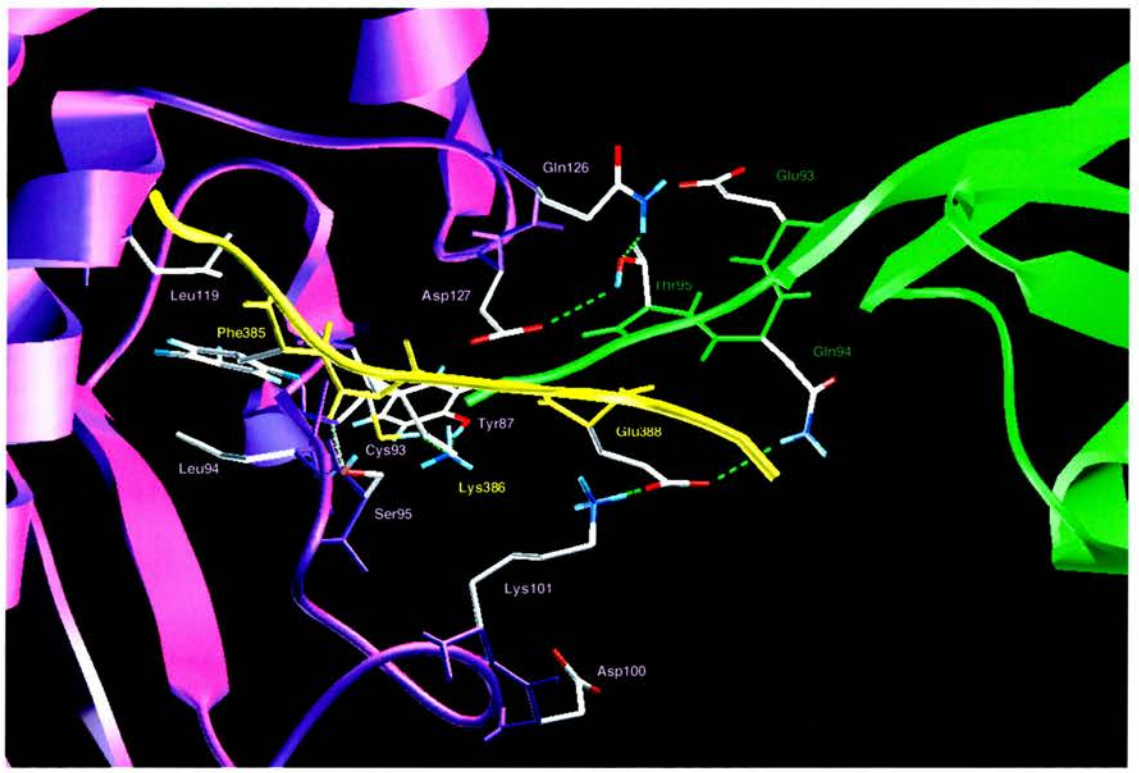


Figure 27. Molecular model of Ubc9, SUMO-1 and a fragment of p53 containing the target lysine. Ribbon representation of a molecular model of a complex between SUMO-1 (green), Ubc9 (magenta) and a fragment of p53 containing the target Lys386 (383-390) (yellow). The model was generated initially by eye, before energy minimisation using GROMOS96 43B1 from the Swiss Pdb viewer software (Guex *et al.*, 1999). N- and C-termini, and proposed regions of E1 and whole substrate interaction shown. Image generated using Swiss PDB viewer package v3.6b (Guex *et al.*, 1999).

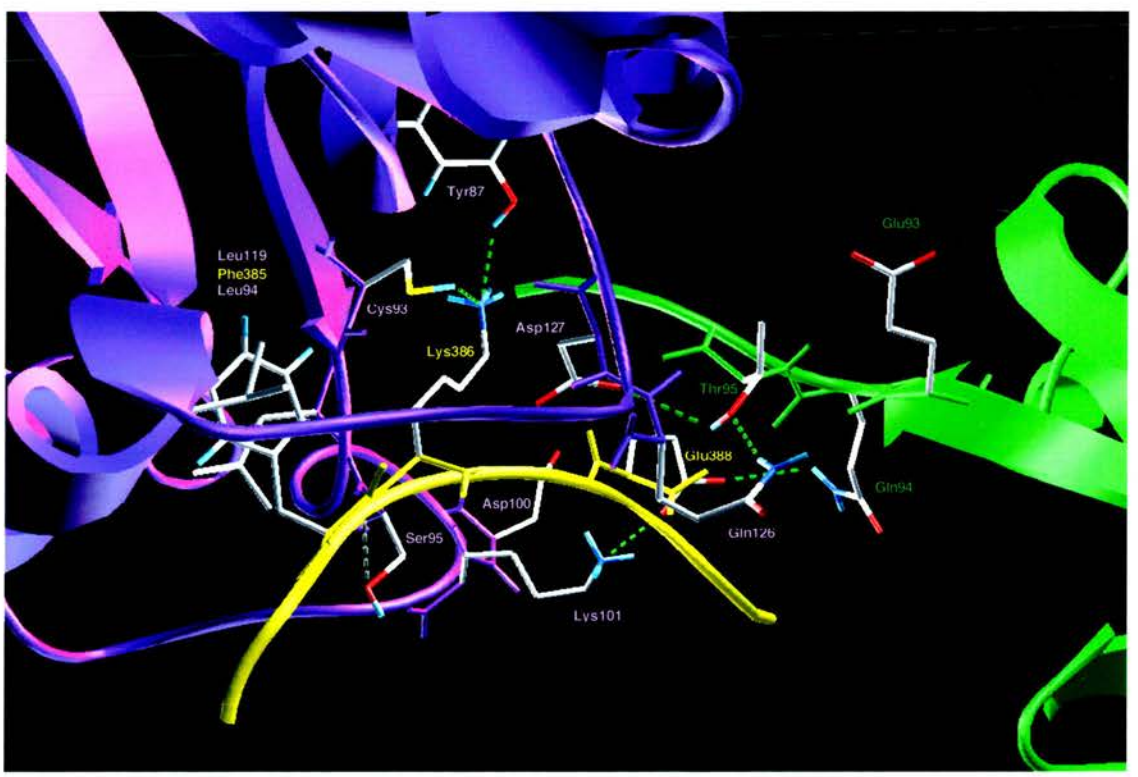
protruding out from the complex perpendicular to Ubc9, it is possible that other contacts are made between the proteins which are difficult to predict accurately with our present limited knowledge (see Figure 27). Paradoxically, the Smt3-Ulp1 crystal structure may actually help to indirectly predict the degree of interaction made between SUMO-1 and Ubc9. The structure shows that in addition to amino-acids close to the C-terminus, residues within in the ubiquitin-core domain of Smt3 make contacts with Ulp1, mostly in the region 64-73 (63-72 in SUMO-1). The fact that mutations of residue 93-95 in SUMO-1 have a greater effect on conjugation reactions than maturation and removal, suggests that fewer contacts (i.e. mostly C-terminal) are made between SUMO-1 and Ubc9 in a conjugation complex than between SUMO-1 and the SUMO-1 protease in a proteolytic complex. Furthermore, if the interactions between E2 and modifier are conserved between the ubiquitin and SUMO-1 systems, then a report studying interactions between ubiquitin and Ubc2b supports these conclusions. Using NMR perturbation studies it was shown that ubiquitin only contacts Ubc2b mostly via the six C-terminal residues and to a lesser degree Lys48 and Gln49 (Miura. T, *et al.* 1999). Making this assumption the major role of the SAE1/SAE2 protein in its complex with E2 may be as a 'bridging' molecule between SUMO-1 and Ubc9, recognising large regions of both proteins, and thus compensating for SUMO-1-Ubc9 contacts, leaving a relatively flexible Ubc9-SUMO-1 thioester after trans-esterification. Consistent with the proposed E1 binding site indicated in Figure 27, Glu93 from SUMO-1, shown to be involved in SAE1/SAE2 interaction, points into this region (Figure 28B). Interestingly, this residue

Figure 28. Zoom views showing the interactions made between SUMO-1, Ubc9 and p53 (383-390) in the model complex. Two views are presented of the model which shows SUMO-1 (backbone and ribbon shown in green) and Ubc9 (backbone and ribbon shown in magenta) and the p53 fragment (backbone and ribbon shown in yellow), and all presented side-chains in standard atomic colours; **A.** Directly zooming into the complex from Figure 27, and **B.** zooming in from above. Glu93 from SUMO-1 thought to be involved in E1 bond formation, but has no effect on E2 interaction, points out from the complex into the region between SUMO-1 and Ubc9 thought to be involved in E1 binding. Gln94 from SUMO-1 known to affect transfer from Ubc9 to substrate is shown forming an H-bond (broken green line) with Glu388 from p53, which also forms an ionic interaction with Lys101 from Ubc9, also thought to be involved in transfer of SUMO-1 to substrate. Thr95 from SUMO-1, proposed to be involved in transfer from E1 to Ubc9 is shown forming H-bonds with Ubc9 via Gln126 and Asp127. The target Lys386 from the p53 fragment appears to interact with Ubc9 via an H-bond with Tyr87. Phe385 of p53 is shown flanked by Lue94 and Lue119 from Ubc9 bonding by hydrophobic interactions. The H-bond between Lys386 and Cys93 is an artefact of the model as the thioester bond between Gly97 and Cys93 has not been simulated. Images generated using Swiss PDB viewer package v3.6b (Guex *et al.*, 1999).

A


Gln126
Glu93
Leu119
Asp127
Thr95
Phe385
Gln94
Tyr87
Glu388
Cys93
Leu94
Lys386
Ser95
Lys101
Asp100

B


Tyr87
Glu93
Leu119
Phe385
Leu94
Cys93
Asp127
Lys386
Thr95
Glu388
Gln94
Asp100
Ser95
Gln126
Lys101

is substituted with glutamine in SUMO-2 and SUMO-3, and as such, any important side-chain interactions are likely to be via the carboxyl oxygen common to both residues.

The ε-amino group of the target Lys386 of p53 is shown to make an H-bond with Tyr 87 of Ubc9 when located close to Cys93 (clearest in Figure 28B). Although the model simulates, but does not actually show the thioester bond formed between the SUMO-1 C-terminal Gly97 and Cys93 of Ubc9 (and thus shows an artefactual H-bond between Cys93 and Lys386), this may still be a feasible interaction in the thioester complex, and so may be an important determinant in locating the lysine side-chain into a catalytically favourable position. The orientation of the p53 fragment is such that Phe385 (Ψ from the Ψ**K**XE consensus motif) sits in a small hydrophobic pocket in Ubc9 created by Leu94 and Leu119 (Figure 28), which is large enough to accommodate any large hydrophobic residue. Also Glu388 (E from the Ψ**K**XE consensus motif) forms an ionic bond with Lys101 from Ubc9 (implicated in substrate recognition), and an H-bond with Gln94 from SUMO-1 (involved in transfer of SUMO-1 from Ubc9 to substrate). As such Ubc9 Lys101 and SUMO-1 Gln94 work together in this model to orient Glu388 of the target protein correctly in the complex. Interestingly in the model the p53 fragment appears to bend down towards the active site cysteine residue (Figure 28B). Although the crystal structure of the entire p53 protein has not been resolved, and thus cannot be mapped onto the model, it is likely that this represents a deformation of the molecule which pulls Lys386 into the catalytic site. Should this change

not occur, it would likely affect catalysis but not overall substrate binding, supporting the observed characteristics of the Asp100-Lys101 loop mutations.

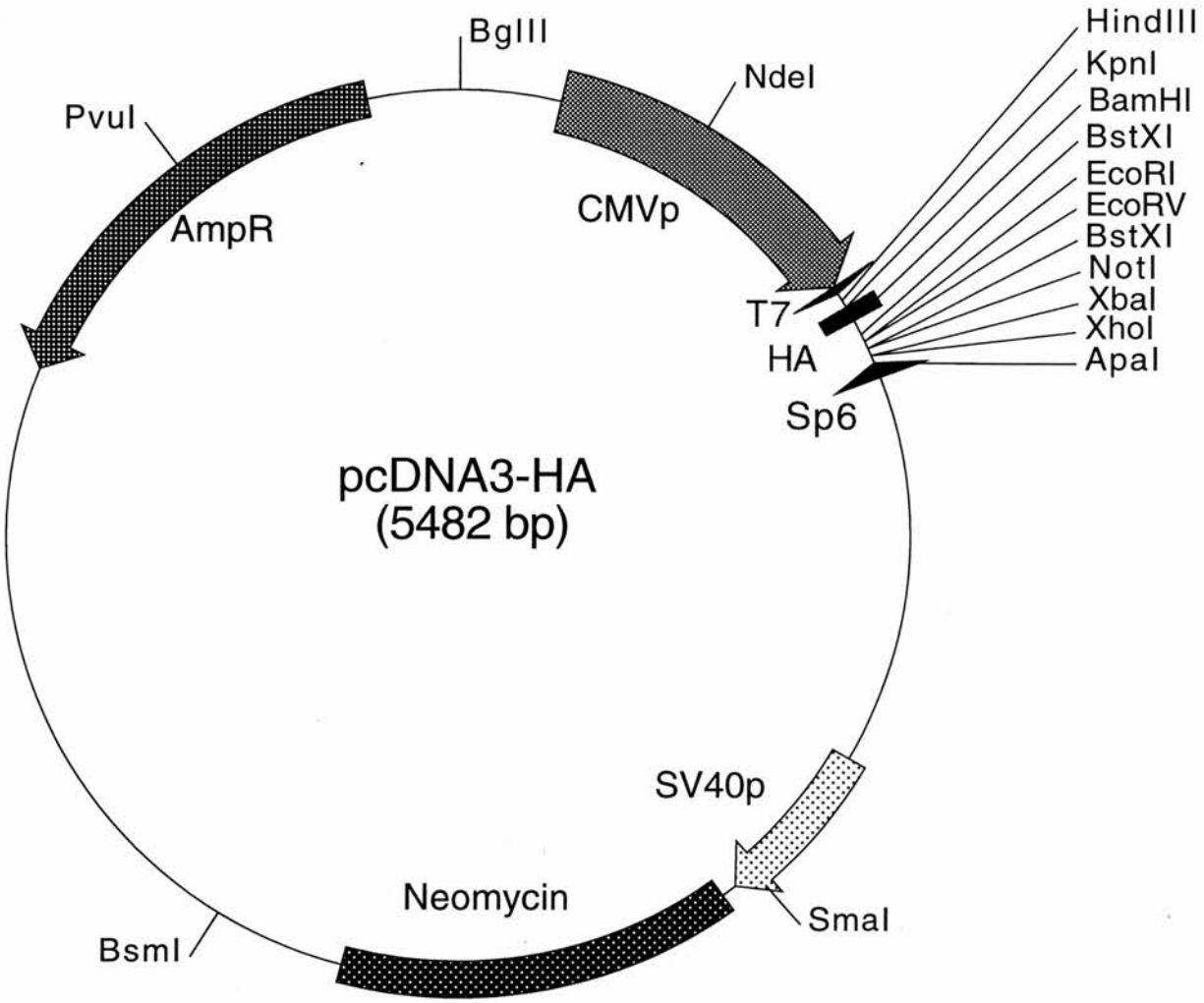
In the model Thr95 from SUMO-1 shown to be important for transfer of SUMO-1 from E1 to Ubc9, forms strong interactions with Ubc9 via two H-bonds with Gln126 and Asp127. Clearly, mutation of Thr95 to Arg would disrupt such associations and may account for the inhibition of the trans-esterification reaction as observed in the Thr95-Arg SUMO-1 mutant.

It is important to note that this model does not account for any major back-bone movements in either SUMO-1 or Ubc9, which may be the explanation for the lack of the predicted involvement of Asp100 in the recognition of the consensus motif. As both Asp100 and Lys101 form part of a large flexible loop close to the active site, it is possible that a conformational change occurs upon substrate binding which brings Asp100 into proximity with the substrate. It is also possible that Asp100 is involved transiently in the complex formation, possibly interacting with the acceptor lysine, but not involved in stabilizing the final conformation. Alternatively it may not alter position significantly, and simply be involved in main chain hydrogen bonding of a region of p53 close to the consensus which is important in locating Lys386 near to Cys93 of Ubc9 for efficient catalysis. The corresponding region to the loop into which Asp100-Lys101 is inserted, and another loop closest to it in a ubiquitin E2, have been shown to be involved in E3 binding in the recently published structure of a UbcH7-c-Cbl complex (Zheng *et al.*, 2000). Thus as Ubc9 does

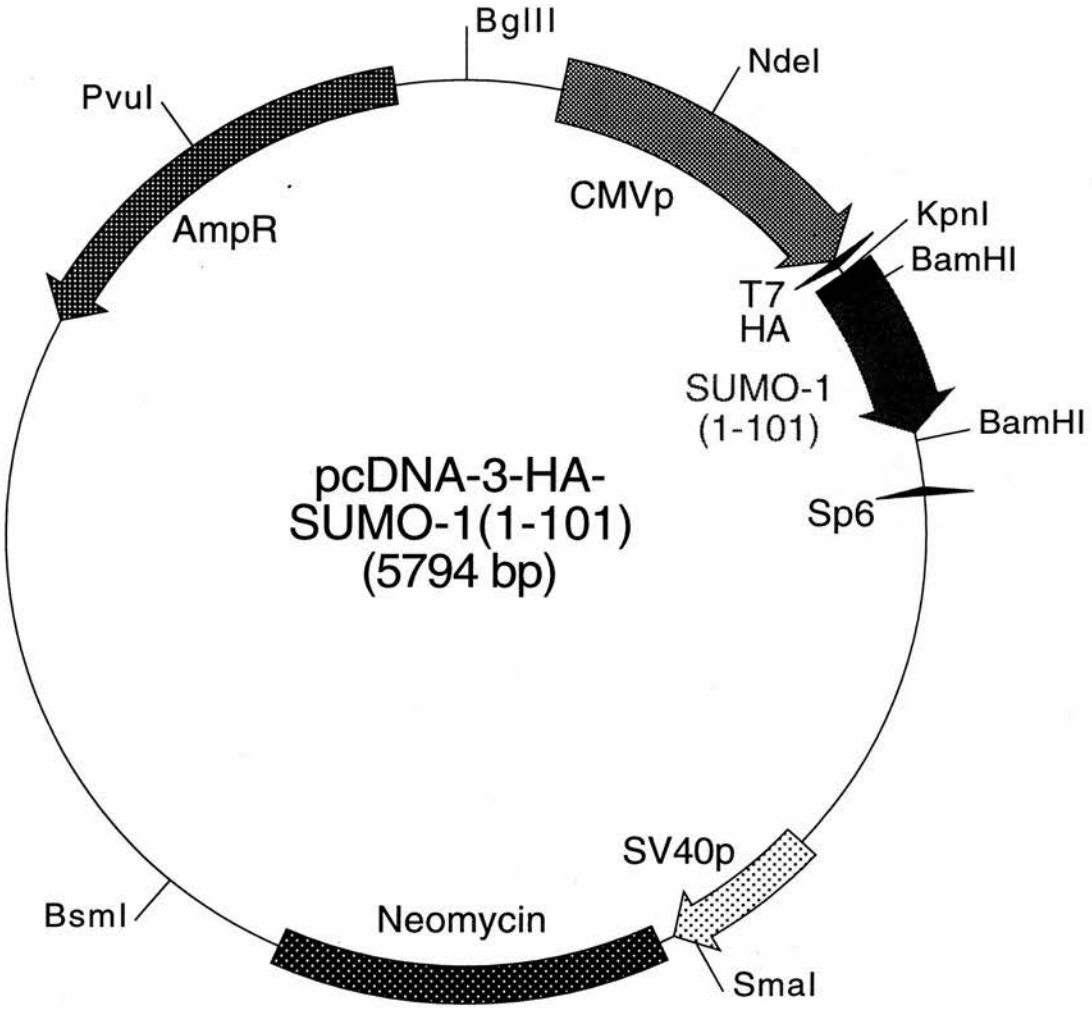
not absolutely require E3 proteins, it is conceivable that these regions of Ubc9 are involved in the major interactions with substrates (shown in Figure 27).

Clearly the model involves a large degree of speculation, however it has revealed a number of possible interactions between SUMO-1, Ubc9 and substrate which may be important for catalysis. Thus it would be interesting to investigate conservative Ubc9 site-directed mutants which would abrogate the proposed interactions such as Gln126-Glu, Asp127-Leu, Tyr87-Phe, Leu94-Asn and Leu119-Asn without compromising overall side-chain structure. However, the ultimate goal of future work would be to resolve the X-ray crystallography structure of a SUMO-1-Ubc9-substrate (fragment) complex. Although likely to represent a challenging project, the success of such an investigation would not only have implications within the field of SUMO-1 modification, but all ubiquitin and ubiquitin-like protein modification pathways.

5. APPENDIX - VECTOR MAPS



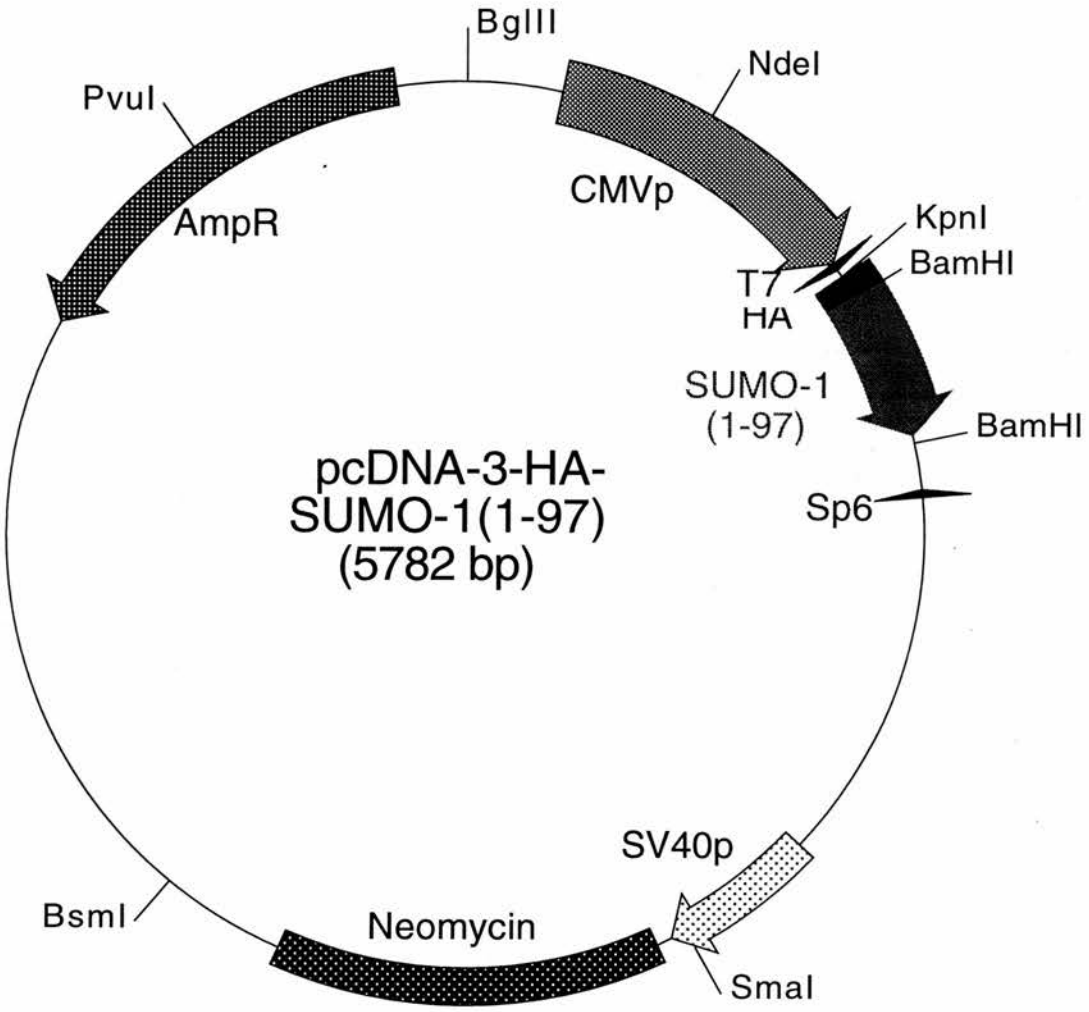
NOTES: pcDNA3-HA (5482 nucleotides)
A gift from J. M. Desterro (Univerity of St. Andrews)
Supplier of pcDNA3: Invotrogen
pcDNA3-HA differs from the parent plasmid by the insertion of the 42 bp sequence ATGGCTTCATATCCTTACGATGTTCCAGACTACGCTTCCCTT between the Kpn1 and BamH1 sites of the polylinker.
CMV promoter: bases 209-863
T7 promoter: bases 864-882
Polylinker: bases 889-1038 (including the HA sequence: bases 900-9042)
Sp6 promoter: bases 1041-1058
BGH polyA: bases 1060-1291
SV40 promoter: bases 1832-2157
SV40 origin of replication: bases 2026-2111
NeoR ORF: bases 2193-2974
SV40 polyA: bases 3162-3292
pUC19 backbone: bases 3314-5488
AAmpR ORF: bases 4492-5352
Antibodies: The HA insert encodes MASYPYDVPDYASL of which YPYDVPDYA is a fragment from the influenza virus haemagglutinin protein recognised by the anti-HA 12CA5 mono-clonal antibody from Babco.



Cloning primers:
Upstream (BamHI): 5'-ACA AAC GGA TCC ATG* TCT GAC CAG GAG GCC AAA-3'
Downstream mutant (BamHI):
RLR-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TA-3'
RR-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TTG TCG CTG ATA AA-3'
T95R-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TTG TTC CTG-3'
Q94L-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TAG TTC CTG ATA AA-3'
E93R-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCG TTT GTC GCT GAT AAA-3'

Organism: *Homo sapiens*
Coding region: 303 bp
Accession numbers of wild-type: NP_003343, Q93068, AAC50996

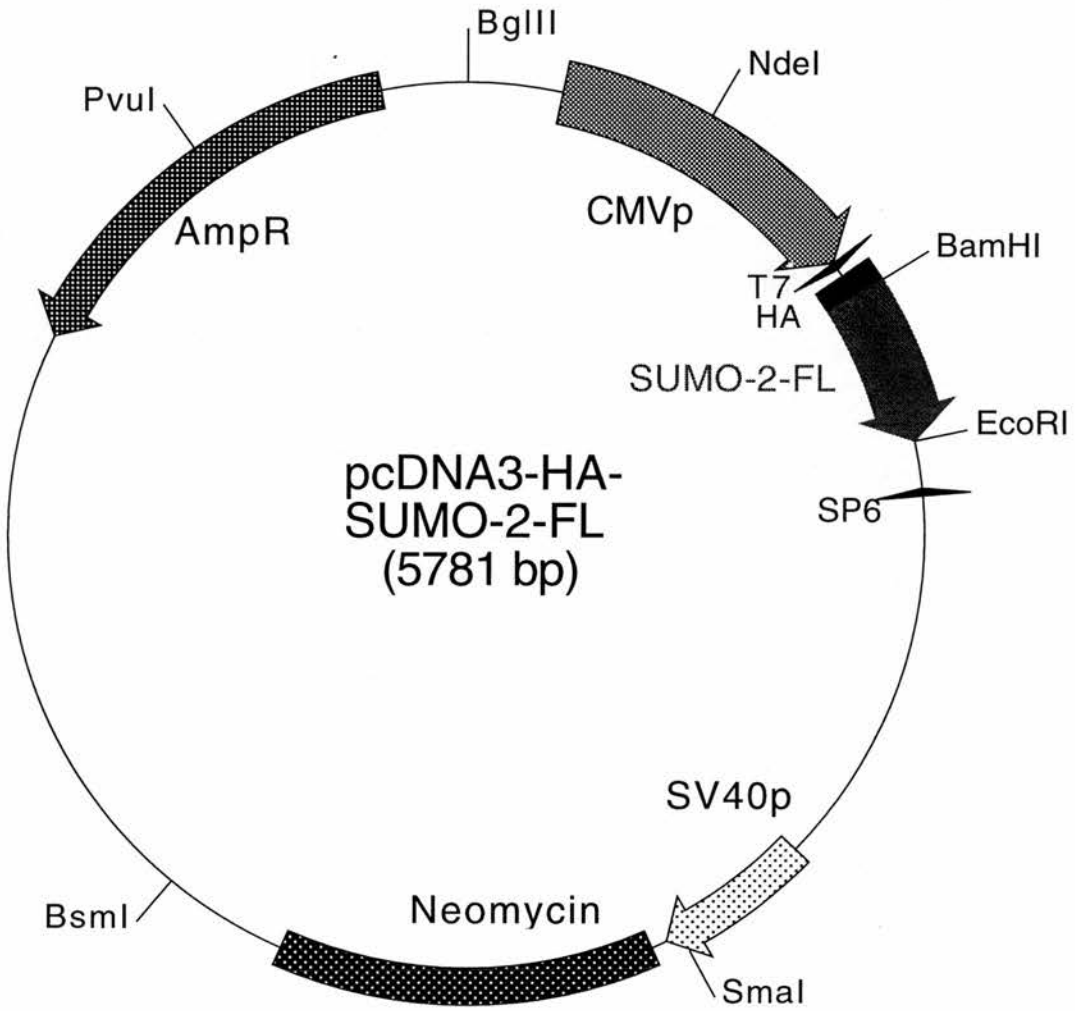
Antibodies:
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).
Mouse 21C7 monoclonal antibody to recombinant GMP1 (SUMO-1) (Zymed-Cambridge Bioscience).



1-97 Cloning primers:
Upstream (BamHI): 5'-ACA AAC GGA TCC ATG* TCT GAC CAG GAG GCC AAA-3'
Downstream (BamHI):
RLR-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CCT TAG TCG CTG ATA AAC-3'
RR-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CCT TTG TCG CTG ATA AAC-3'
T95R-SUMO-1(1-97); 5'-ACA AAC CAT ATG* TCT GAC CAA GGC AAA A-3'
Q94L-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CGT TAG TTC CTG ATA AAC-3'
E93R-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CGT TTG TCG CTG ATA AAC-3'

Organism: *Homo sapiens*
Coding region: 292 bp
Accession numbers of wild-type: NP_003343, Q93068, AAC50996

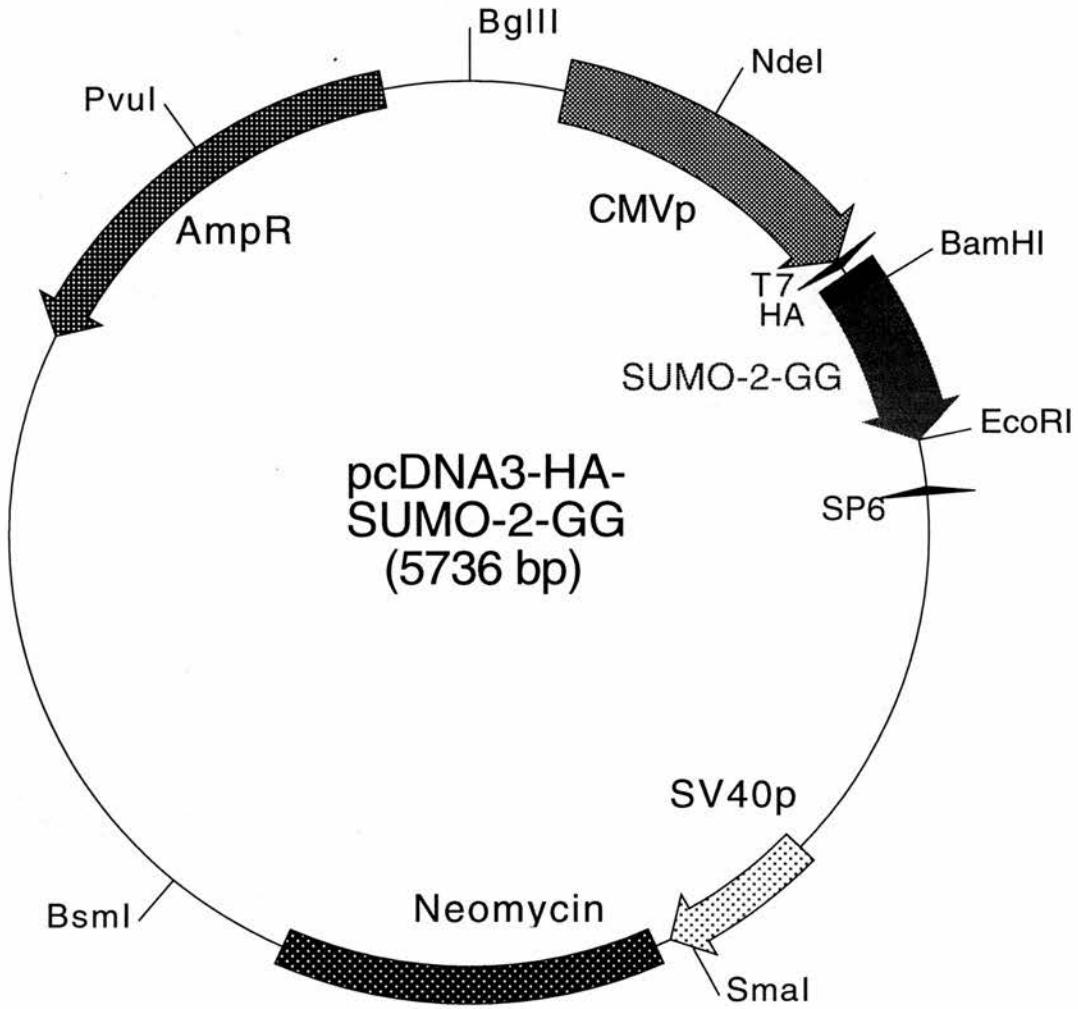
Antibodies:
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).
Mouse 21C7 monoclonal antibody to recombinant GMP1 (SUMO-1) (Zymed-Cambridge Bioscience).



Cloning primers:
Upstream primer (BamHI); 5'-TCC CCG CGC CGC TCG GGA TCC ATG* TCC GAG-3'
Downstream primer (EcoRI); 5'-CCC GAA TTC GGG ACG GGC CCT *CTA GAA ACT-3'

Organism: *Homo sapiens*
Coding region: 309 bp
Accession numbers: P55854, CAA67896.

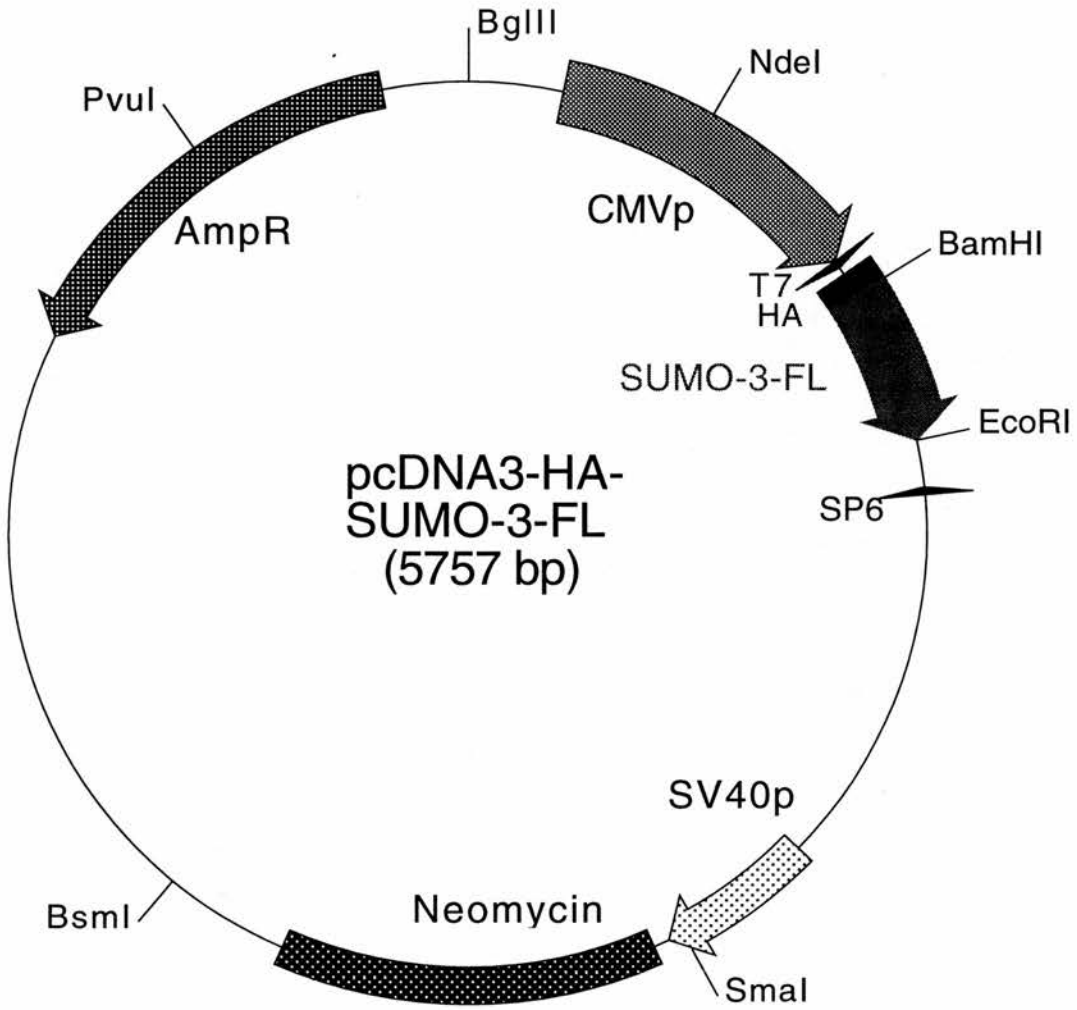
Antibodies:
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).



Cloning primers:
Upstream primer (BamHI); 5'-TCC CCG CGC CGC TCG GGA TCC ATG* TCC GAG-3'
Downstream primer (EcoRI); 5'-CCC GAA TTC *CTA ACC TCC CTG CTG CTG TTG GAA CAC-3'

Organism: *Homo sapiens*
Coding region: 276 bp
Accession numbers: P55854, CAA67896.

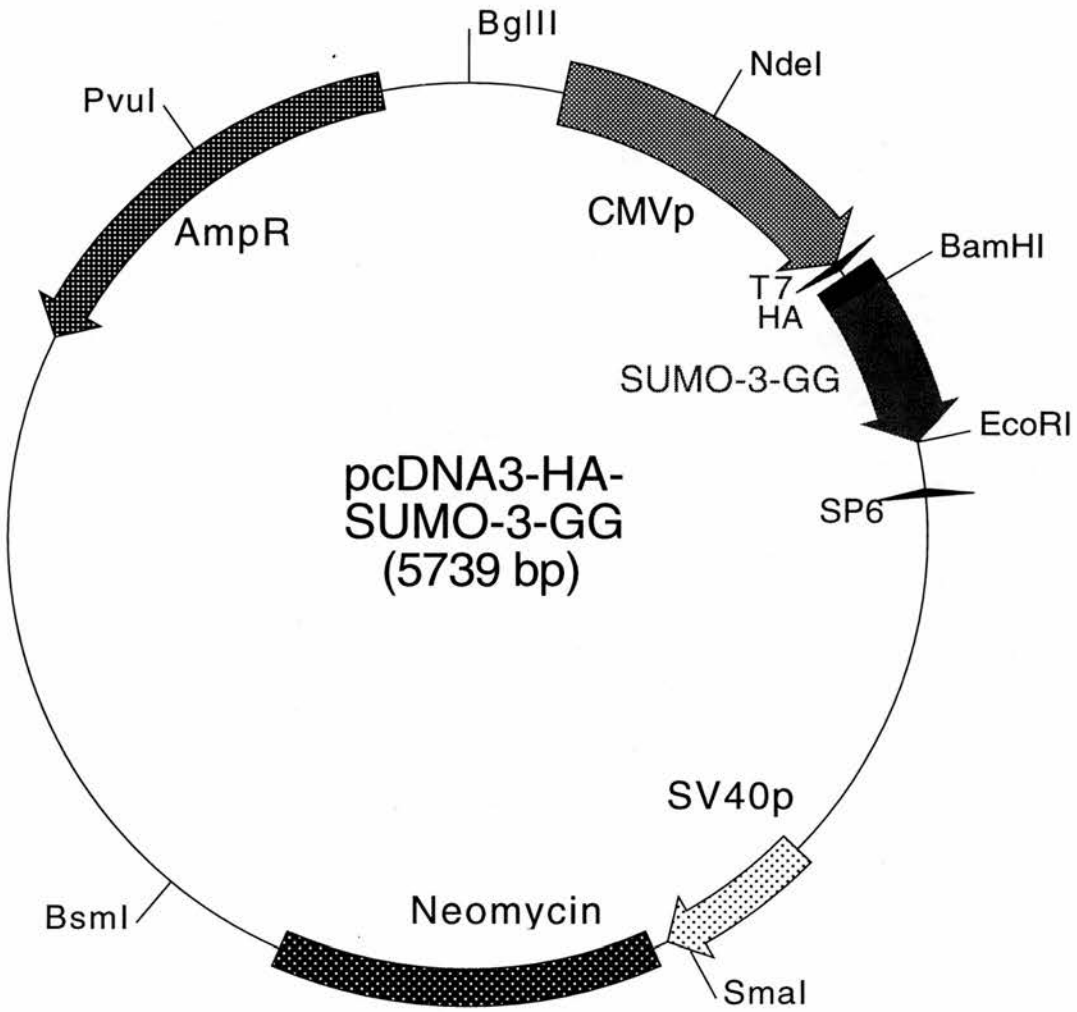
Antibodies:
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).



Cloning primers:
Upstream (BamHI); 5'-GAG GAG ACT CCG GCG GGA TCC ATG* GCC GAC GAA-3'
Downstream (EcoRI); 5'-GTA GAA TTC CAG GTT CCC TTT *TCA GTA GAC-3'

Organism: *Homo sapiens*
Coding region: 285 bp
Accession numbers: P55855, CAA67897.

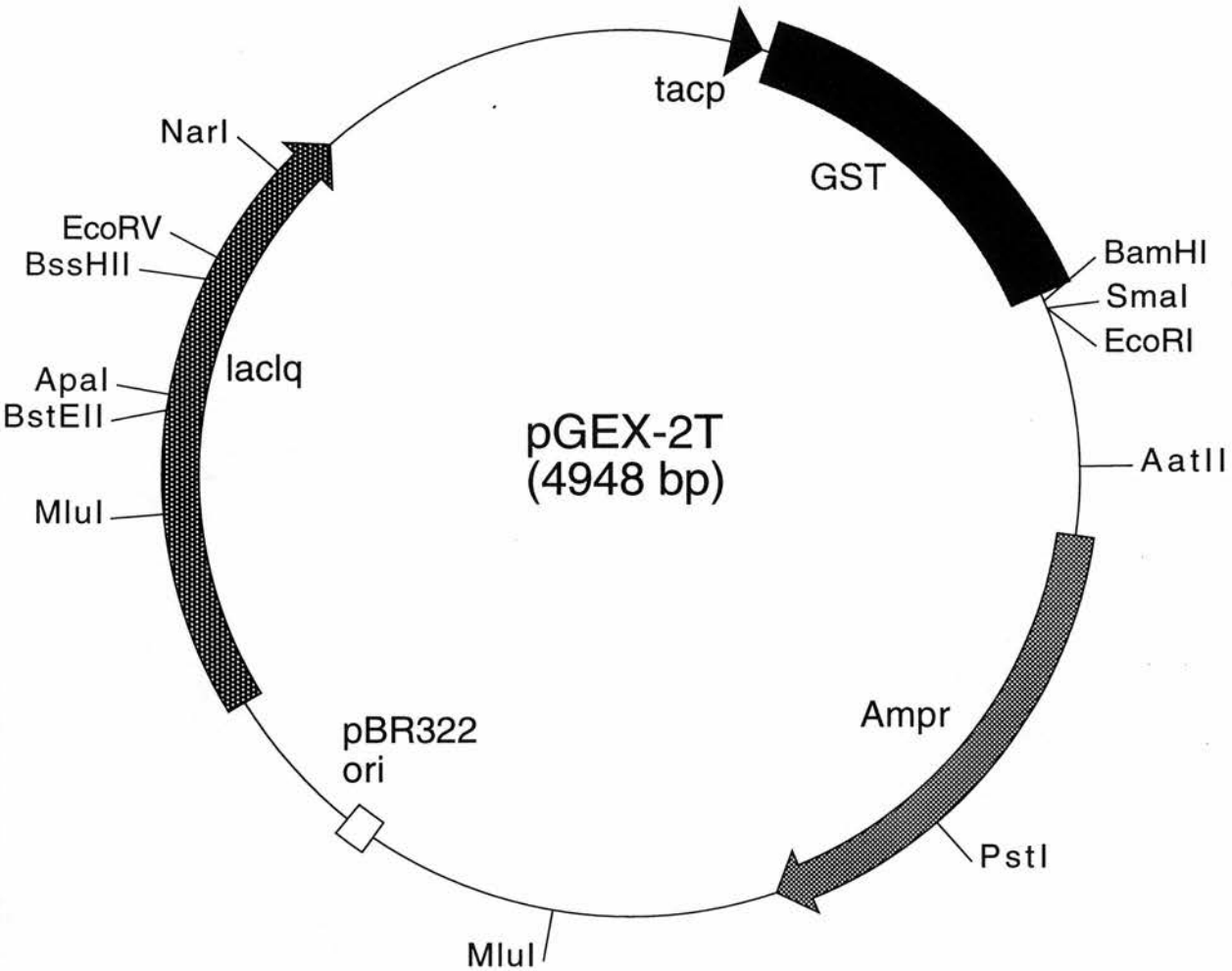
Antibodies:
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).



Cloning primers:
Upstream (BamHI); 5'-GAG GAG ACT CCG GCG GGA TCC ATG* GCC GAC GAA-3'
Downstream (EcoRI); 5'-CCC GAA TTC *CTA ACC TCC CTG CTG CTG TTG GAA CAC-3'

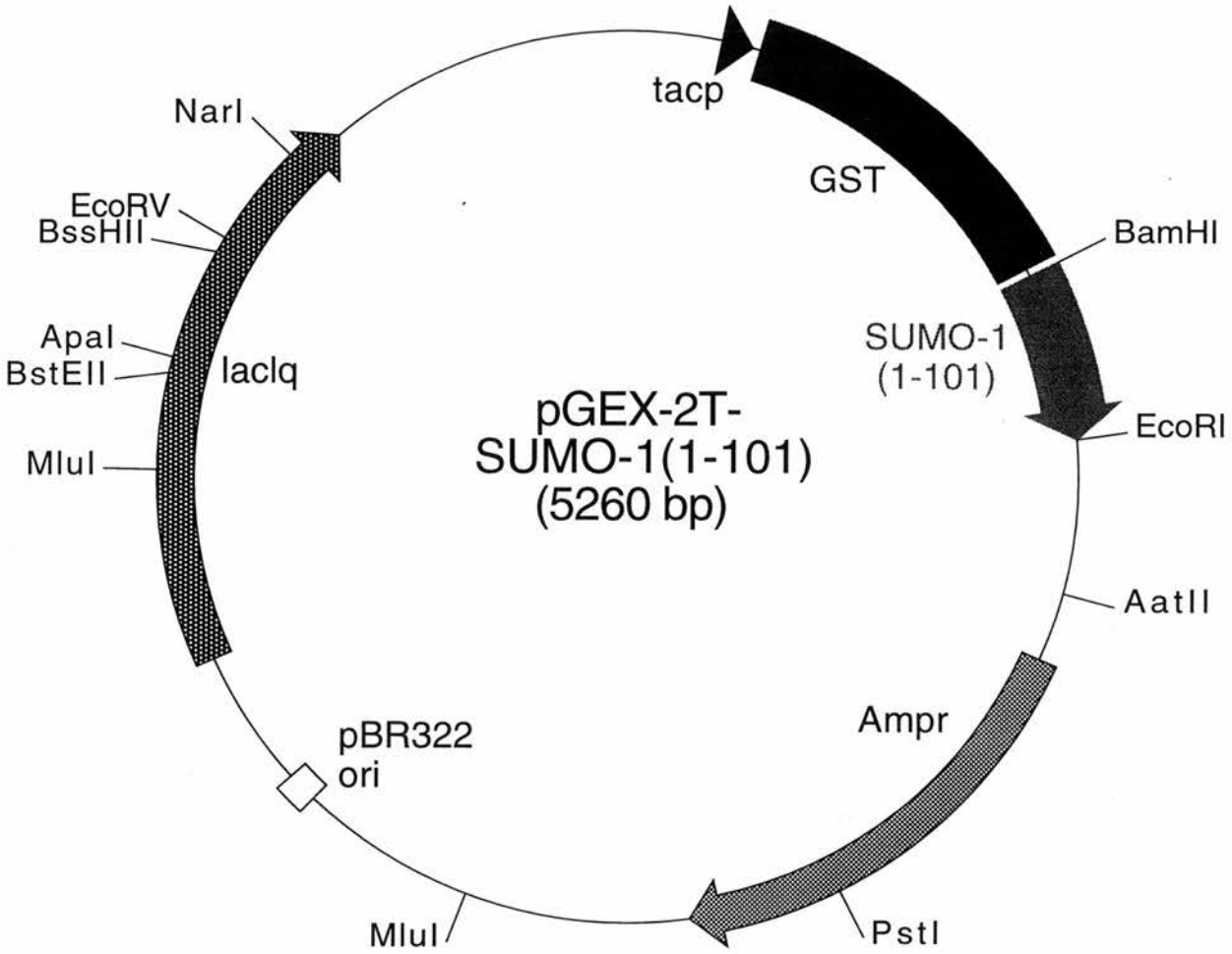
Organism: *Homo sapiens*
Coding region: 279 bp
Accession numbers: P55855, CAA67897.

Antibodies:
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).



Notes: pGEX-2T (4948 nucleotides)
Supplier: Pharmacia
Accession number U13850
tac promoter-35: bases 183-188
tac promoter-10 bases 205-211
lac operator: bases 217-237
GST satrt codon: base 258
Ampr ORF: bases 1356-2214
laclq ORF: bases 3297-4377
Site of replication of initiation: base 2974
Multiple coloning site: bases 930-945

For inducible high level expression of genes as fusions with glutathione-S-transferase (GST) gene from *Schitssoma japonicum,* which forms an affinity tail on the protein products of genes inserted into the multiple cloning site (MCS). Expression is under the control of the tac promoter.

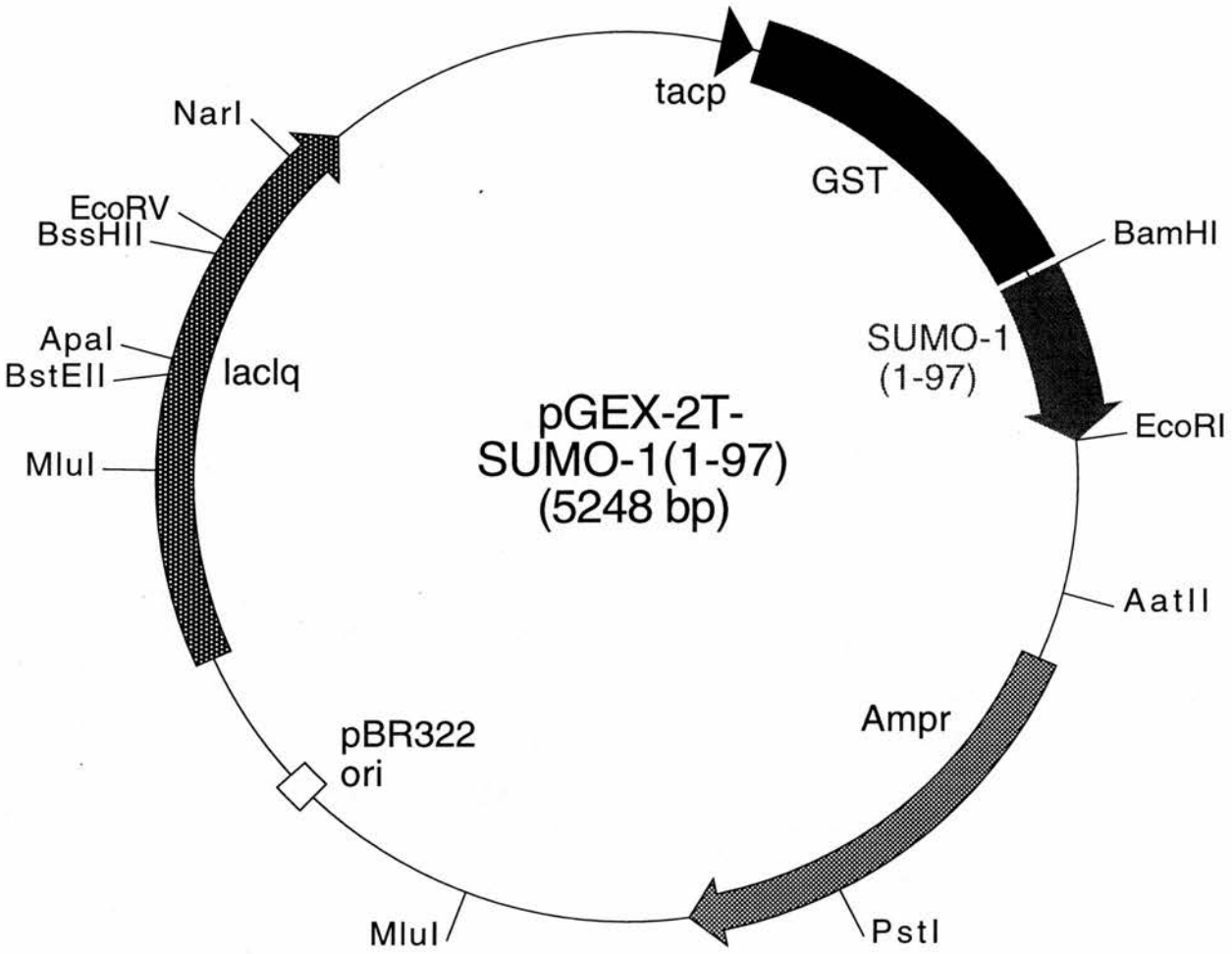


Cloning primers:
Upstream (BamHI): 5'-ACA AAC GGA TCC ATG* TCT GAC CAG GAG GCC AAA-3'
Downstream mutant (BamHI):
RLR-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TA-3'
RR-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TTG TCG CTG ATA AA-3'
T95R-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TTG TTC CTG-3'
Q94L-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCT TAG TTC CTG ATA AA-3'
E93R-SUMO-1(1-101); 5'-ATA ATA GGA TCC *TCA AAC TGT TGA ATG ACC CCC CCG TTT GTC GCT GAT AAA-3'

Organism: *Homo sapiens*
Coding region: 303 bp
Accession numbers of wild-type: NP_003343, Q93068, AAC50996

Bacterial expression:
Expressed in *E. Coli* B834 as GST-fusion proteins.

Antibodies:
Mouse 21C7 monoclonal antibody to recombinant GMP1 (SUMO-1) (Zymed-Cambridge Bioscience).

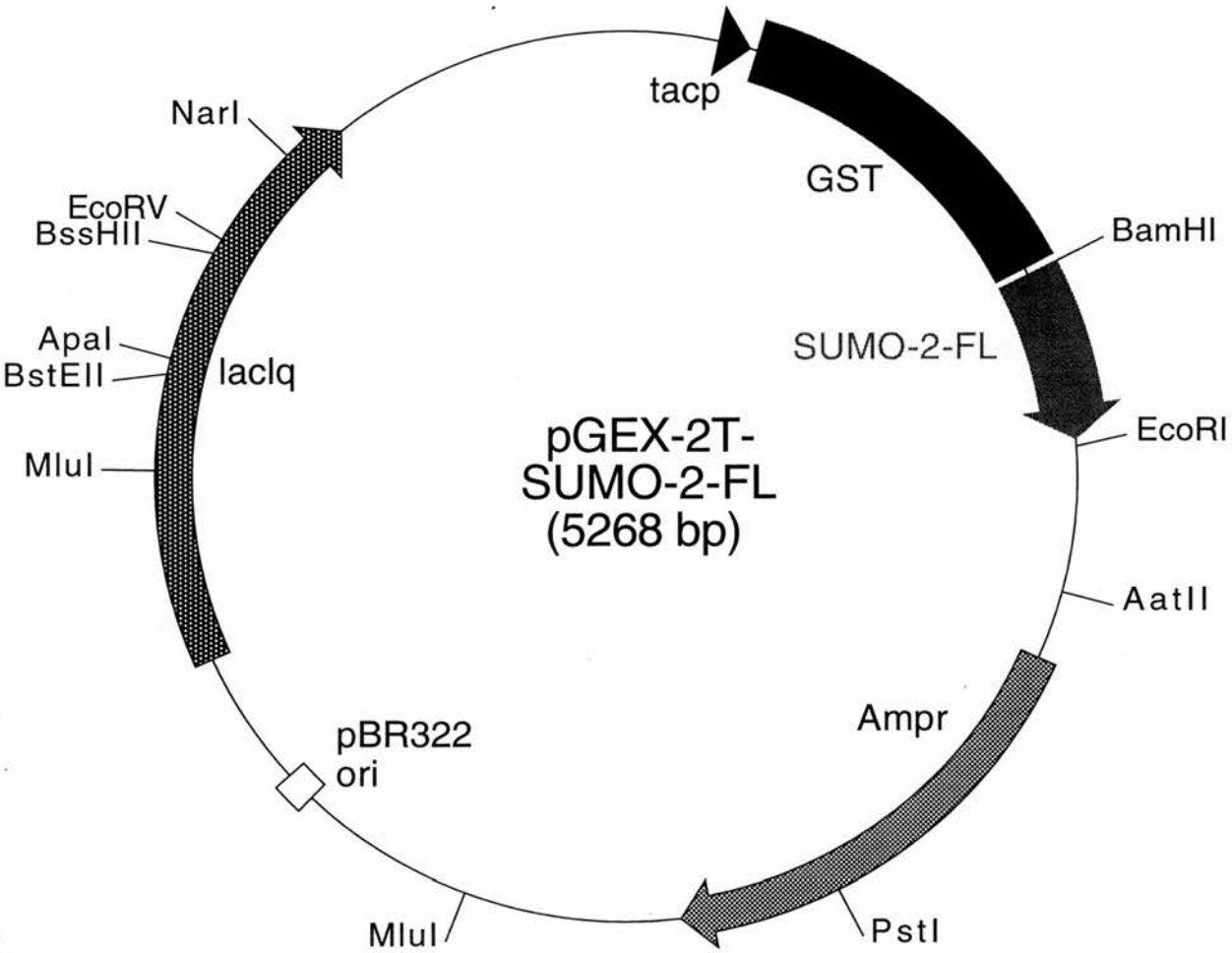


1-97 Cloning primers:
Upstream (BamHI): 5'-ACA AAC GGA TCC ATG* TCT GAC CAG GAG GCC AAA-3'
Downstream (BamHI):
RLR-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CCT TAG TCG CTG ATA AAC-3'
RR-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CCT TTG TCG CTG ATA AAC-3'
T95R-SUMO-1(1-97); 5'-ACA AAC CAT ATG* TCT GAC CAA GGC AAA A-3'
Q94L-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CGT TAG TTC CTG ATA AAC-3'
E93R-SUMO-1(1-97); 5'-ATA ATA GGA TCC *TCA ACC CCC CGT TTG TCG CTG ATA AAC-3'

Organism: *Homo sapiens*
Coding region: 292 bp
Accession numbers of wild-type: NP_003343, Q93068, AAC50996

Bacterial expression:
Expressed in *E. Coli* B834 or BL21 as GST-fusion proteins.

Antibodies:
Mouse 21C7 monoclonal antibody to recombinant GMP1 (SUMO-1) (Zymed-Cambridge Bioscience).

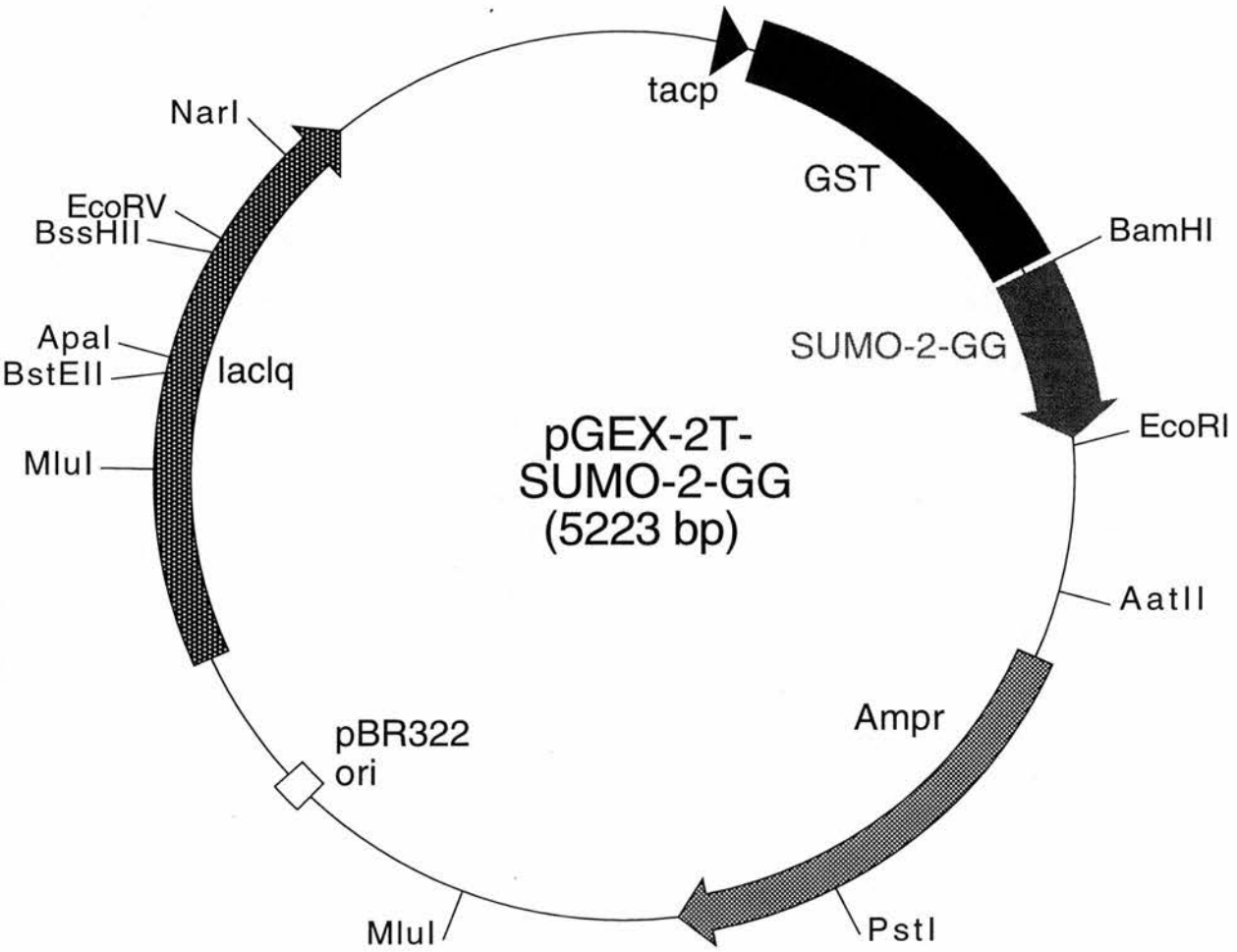


Cloning primers:
Upstream primer (BamHI); 5'-TCC CCG CGC CGC TCG GGA TCC ATG* TCC GAG-3'
Downstream primer (EcoRI); 5'-CCC GAA TTC GGG ACG GGC CCT *CTA GAA ACT-3'

Organism: *Homo sapiens*
Coding region: 309 bp
Accession numbers: P55854, CAA67896.

Bacterial expression:
Expressed in *E. Coli* B834 as GST-fusion proteins.

Antibodies:
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).

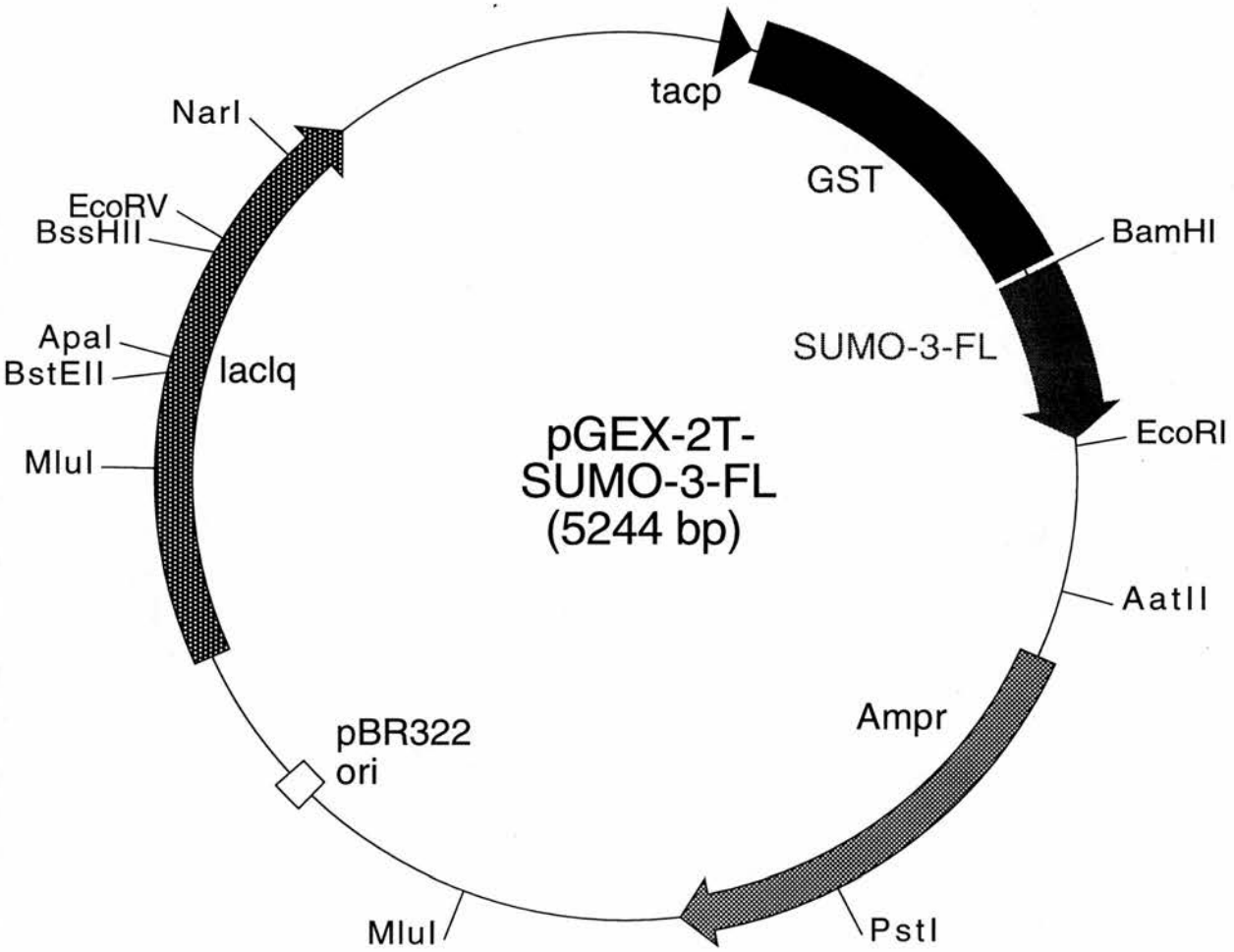


Cloning primers:
Upstream primer (BamHI); 5'-TCC CCG CGC CGC TCG GGA TCC ATG* TCC GAG-3'
Downstream primer (EcoRI); 5'-CCC GAA TTC *CTA ACC TCC CTG CTG CTG TTG GAA CAC-3'

Organism: *Homo sapiens*
Coding region: 276 bp
Accession numbers: P55854, CAA67896.

Bacterial expression:
Expressed in *E. Coli* B834 as GST-fusion proteins.

Antibodies:
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).

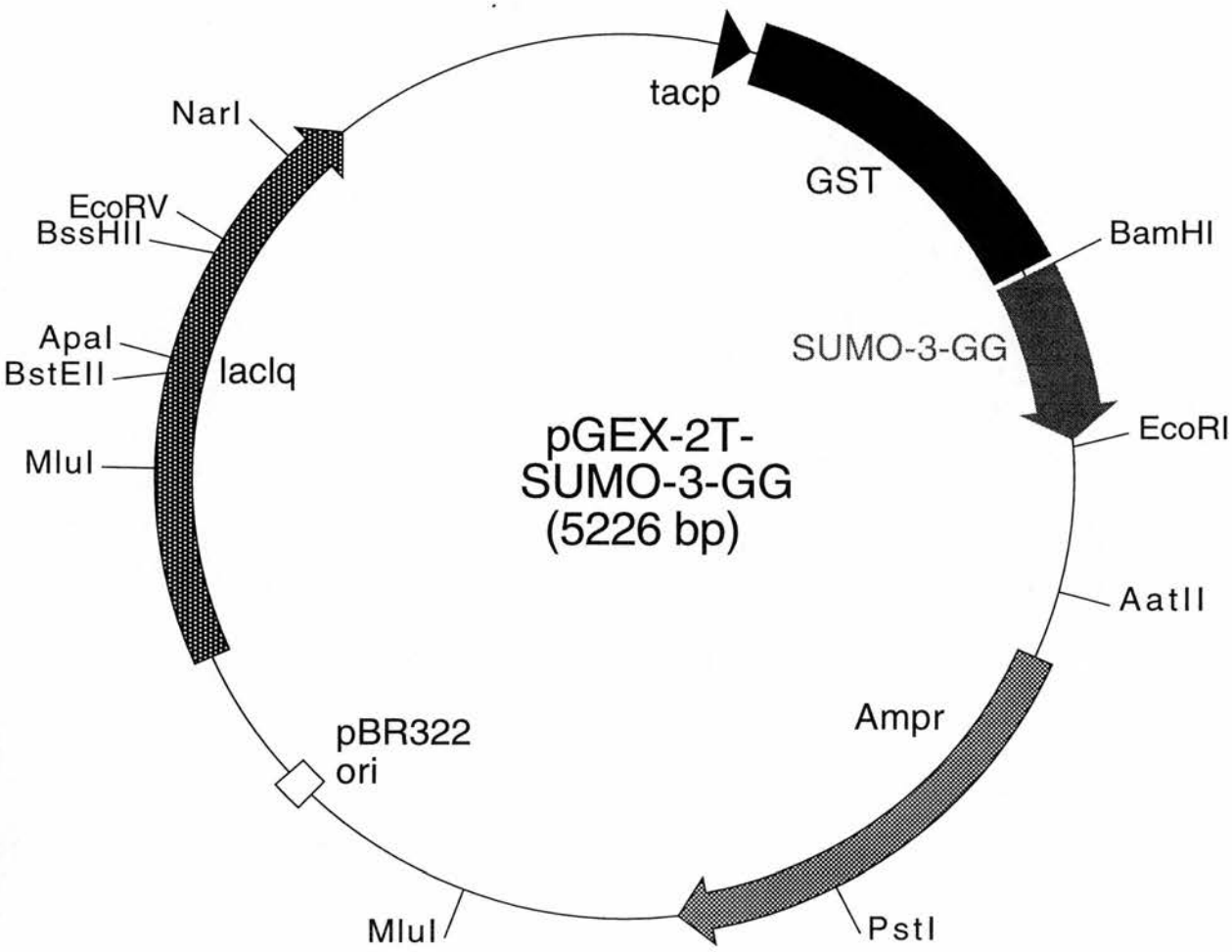


Cloning primers:
Upstream (BamHI); 5'-GAG GAG ACT CCG GCG GGA TCC ATG* GCC GAC GAA-3'
Downstream (EcoRI); 5'-GTA GAA TTC CAG GTT CCC TTT *TCA GTA GAC-3'

Organism: *Homo sapiens*
Coding region: 285 bp
Accession numbers: P55855, CAA67897.

Bacterial expression:
Expressed in *E. Coli* B834 as GST-fusion proteins.

Antibodies:
Mouse 12CA5 monoclonal antibody against HA peptide (Babco).
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).

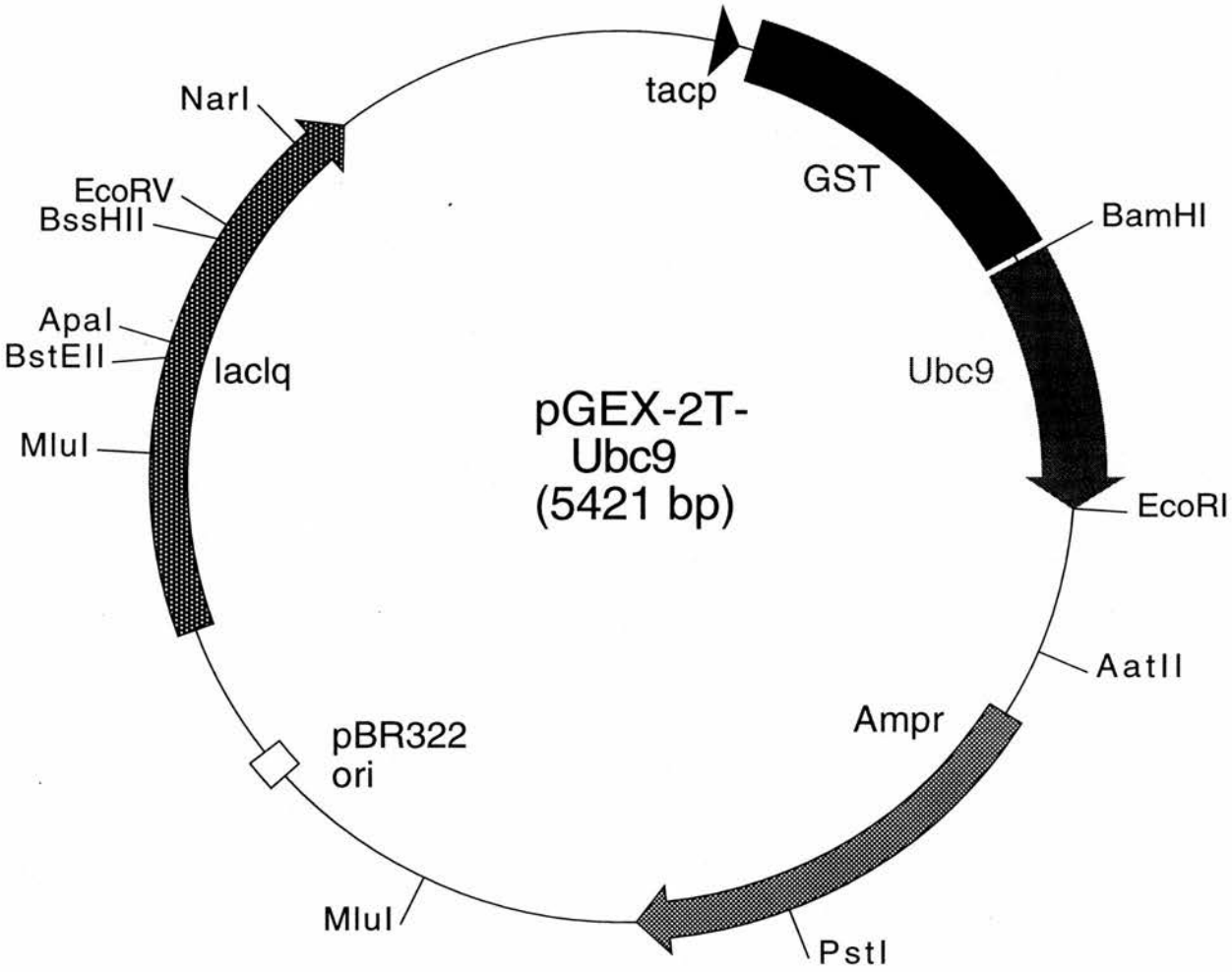


Cloning primers:
Upstream (BamHI); 5'-GAG GAG ACT CCG GCG GGA TCC ATG* GCC GAC GAA-3'
Downstream (EcoRI); 5'-CCC GAA TTC *CTA ACC TCC CTG CTG CTG TTG GAA CAC-3'

Organism: *Homo sapiens*
Coding region: 279 bp
Accession numbers: P55855, CAA67897.

Bacterial expression:
Expressed in *E. Coli* B834 as GST-fusion proteins.

Antibodies:
Rabbit anti SUMO-2/-3 pAB (H.Saitoh Picower Institute, New York).

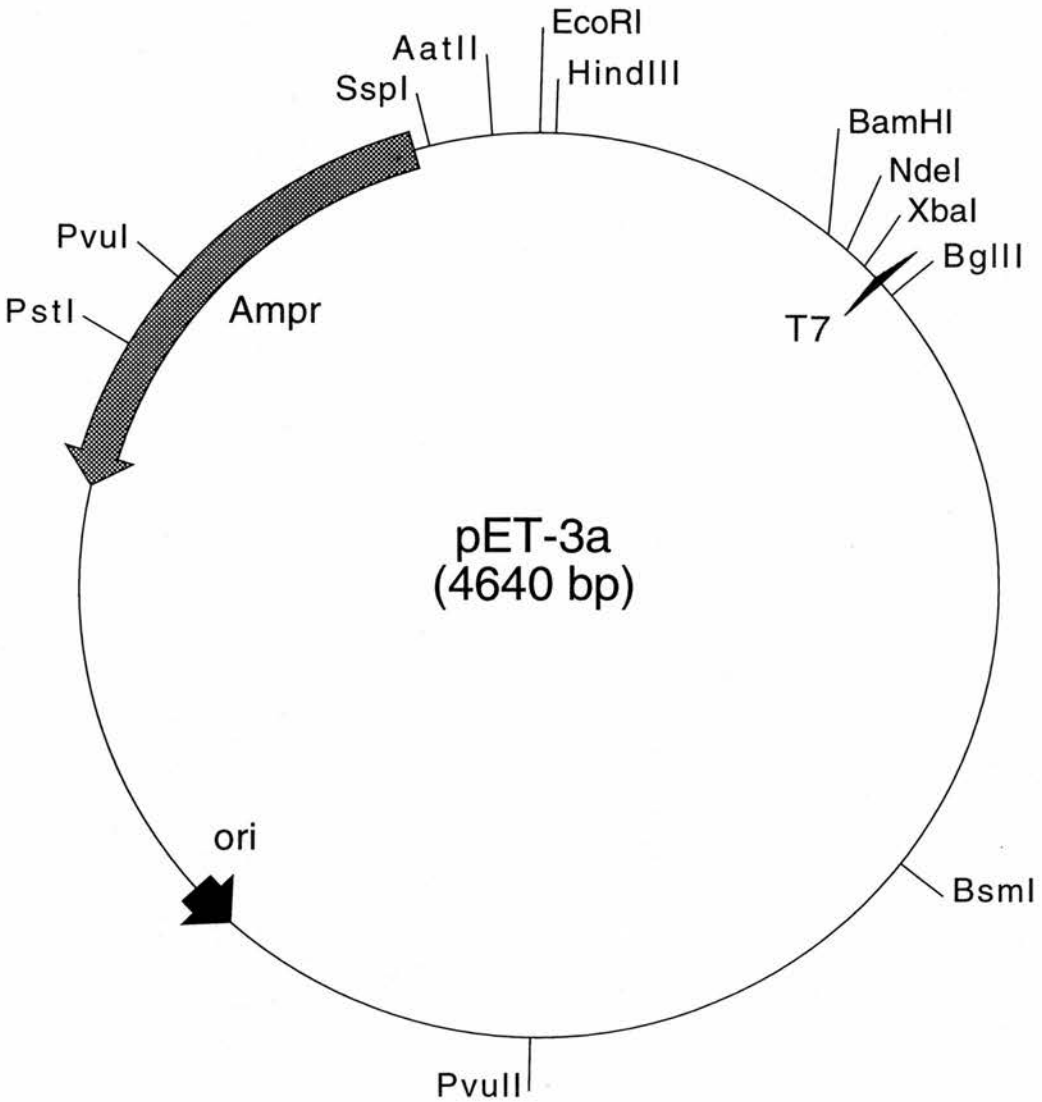


Cloning primers:
External upstream (BamHI); 5'-ACA AAC GGA TCC ATG* TCG GGG ATC GCC CTC AGC-3'
External downstream (EcoRI); 5'-GCC GCG GAA TTC *TTA TGA GGG CGC AAA CTT CTT GGC-3'.
Mutant internal:
Upstream D100A-K101A-Ubc9 (DK-AA-Ubc9); 5'-CCT CCA GTC CGC GGC CTC CTC TAA-3'
Downstream D100A-K101A-Ubc9 (DK-AA-Ubc9); 5'-TTA GAG GAG GCC GCG GAC TGG AGG-3'
Upstream D100A-Ubc9; 5'-CCA GTC CTT GGC CTC CTC TAA-3'
Downstream D100A-Ubc9; 5'-TTA GAG GAG GCC AAG GAC TGG-3'
Upstream K101A-Ubc9; 5'-CCT CCA GTC CGC GTC CTC CTC-3'
Downstream K101A-Ubc9; 5'-GAG GAG GAC GCG GAC TGG AGG-3'

Organism: *Homo sapiens*
Coding region: 474 bp
Accession numbers of wild-type: NM_003345, U29092, U31882, AJ002385, U31933, X96427.

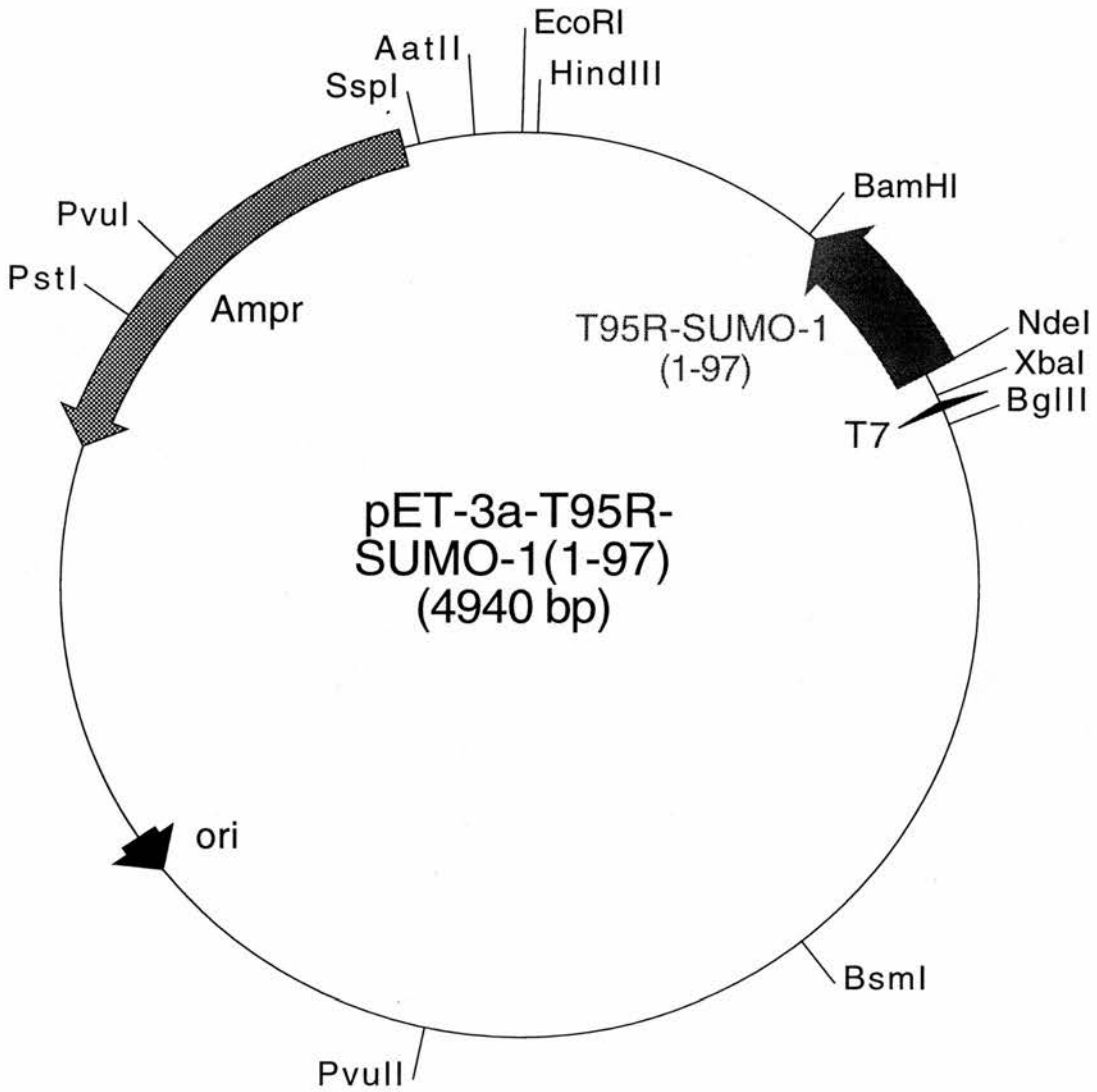
Bacterial expression:
Expressed in *E. Coli* DH-5a as GST-fusion protein.

Antibodies:
Sheep anti-Ubc9 polyclonal antibody against Ubc9 C-terminal peptide CEYEKRVRAQAKKFAPS.



Notes: pET-3a (4640 nucleotides)
Supplier: Novagen
T7 promoter: bases 615-631
T7 transcription start: base 614
T7-tag coding sequence: bases 519-551
T7 terminator: bases 404-450
pBR322 origin: base 2814
bla coding sequence: bases 3575-4432

For the inducible expression of high levels of protein, under the control of the bacteriophage T7 transcription and translation signals. T7 RNApolymerase is sourced from within the host cells.



Cloning primers:
Upstream (NdeI); 5'-ACA AAC CAT ATG* TCT GAC CAA GGC AAA A-3'
Downstream mutant (NdeI); 5'-ATA ATA GGA TCC *TCA ACC CCC CCT TTG TTC CTG -3'

Organism: *Homo sapiens*
Coding region: 292 bp
Accession numbers of wild-type: NP_003343, Q93068, AAC50996

Bacterial expression:
Expressed in *E. Coli* BL21-de3.

Antibodies:
Mouse 21C7 monoclonal antibody to recombinant GMP1 (SUMO-1) (Zymed-Cambridge Bioscience).

6. BIBLIOGRAPHY

Aletta, J.M., Cimato, T.R. and Ettinger, M.J. (1998) Protein methylation: a signal event in post-translational modification. *Trends Biochem Sci*, 89-91.

Allfrey, V., Faulkner, R.M. and Mirsky, A.E. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci*, 786-794.

Apweiler, R., Hermjakob, H. and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta*, 4-8.

Archer, S.Y. and Hodin, R.A. (1999) Histone acetylation and cancer. *Curr Opin Genet Dev*, 171-174.

Arenzana-Seisdedos, F., Turpin, P., Rodriguez, M., Thomas, D., Hay, R.T., Virelizier, J.L. and Dargemont, C. (1997) Nuclear localization of I kappa B alpha promotes active transport of NF-kappa B from the nucleus to the cytoplasm. *J Cell Sci*, 369-378.

Arnason, T. and Ellison, M.J. (1994) Stress resistance in Saccharomyces cerevisiae is strongly correlated with assembly of a novel type of multiubiquitin chain. *Mol Cell Biol*, 7876-7883.

Barik, S. (1996) Protein phosphorylation and signal transduction. *Subcell Biochem*, 115-164.

Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R. and Becker, J. (1998) Structure determination of the small ubiquitin-related modifier SUMO-1. *J Mol Biol*, 275-286.

Bayle, J.H. and Crabtree, G.R. (1997) Protein acetylation: more than chromatin modification to regulate transcription. *Chem Biol*, 885-888.

Berthiaume, L. and Resh, M.D. (1995) Biochemical characterization of a palmitoyl acyltransferase activity that palmitoylates myristoylated proteins. *J Biol Chem*, 22399-22405.

Bhaskar, V., Valentine, S.A. and Courey, A.J. (2000) A functional interaction between dorsal and components of the Smt3 conjugation machinery. *J Biol Chem*, 4033-4040.

Biggins, S., Ivanovska, I. and Rose, M.D. (1996) Yeast ubiquitin-like genes are involved in duplication of the microtubule organizing center. *J Cell Biol*, 1331-1346.

Bischoff, F.R., Krebber, H., Kempf, T., Hermes, I. and Ponstingl, H. (1995) Human RanGTPase-activating protein RanGAP1 is a homologue of yeast

Rna1p involved in mRNA processing and transport. *Proc Natl Acad Sci U S A*, 1749-1753.

Bizzozero, O.A. (1997) The mechanism and functional roles of protein palmitoylation in the nervous system. *Neuropediatrics*, 23-26.

Boddy, M.N., Howe, K., Etkin, L.D., Solomon, E. and Freemont, P.S. (1996) PIC 1, a novel ubiquitin-like protein which interacts with the PML component of a multiprotein complex that is disrupted in acute promyelocytic leukaemia. *Oncogene*, 971-982.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 248-254.

Brunger, A.T. (1993) X-PLOR. Howard Hughes Medical Institute and Yale University, New Haven CT.

Brunger, A.T. (1996) X-PLOR. Howard Hughes Medical Institute and Yale University., New Haven CT.

Buee, L., Bussiere, T., Buee-Scherrer, V., Delacourte, A. and Hof, P.R. (2000) Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Brain Res Rev*, 95-130.

Burch, T.J. and Haas, A.L. (1994) Site-directed mutagenesis of ubiquitin. Differential roles for arginine in the interaction with ubiquitin-activating enzyme. *Biochemistry*, 7300-7308.

Buschmann, T., Fuchs, S.Y., Lee, C.G., Pan, Z.Q. and Ronai, Z. (2000) SUMO-1 modification of Mdm2 prevents its self-ubiquitination and increases Mdm2 ability to ubiquitinate p53. *Cell*, 753-762.

Camp, L.A., Verkruyse, L.A., Afendis, S.J., Slaughter, C.A. and Hofmann, S.L. (1994) Molecular cloning and expression of palmitoyl-protein thioesterase. *J Biol Chem*, 23212-23219.

Casey, P.J. (1995) Protein lipidation in cell signaling. *Science*, 221-225.

Chakrabarti, S.R., Sood, R., Ganguly, S., Bohlander, S., Shen, Z. and Nucifora, G. (1999) Modulation of TEL transcription activity by interaction with the ubiquitin-conjugating enzyme UBC9. *Proc Natl Acad Sci U S A*, 7467-7472.

Chakrabarti, S.R., Sood, R., Nandi, S. and Nucifora, G. (2000) Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell-cycle-dependent assembly into nuclear bodies. *Proc Natl Acad Sci U S A*, 13281-13285.

Chou, T.Y., Hart, G.W. and Dang, C.V. (1995) c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J Biol Chem*, 18961-18965.

Chow, N., Korenberg, J.R., Chen, X.N. and Neve, R.L. (1996) APP-BP1, a novel protein that binds to the carboxyl-terminal region of the amyloid precursor protein. *J Biol Chem*, 11339-11346.

Ciechanover, A. and Schwartz, A.L. (1998) The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death. *Proc Natl Acad Sci U S A*, 2727-2730.

Cohen, P.T., Brewis, N.D., Hughes, V. and Mann, D.J. (1990) Protein serine/threonine phosphatases; an expanding family. *FEBS Lett*, 355-359.

D'Andrea, A. and Pellman, D. (1998) Deubiquitinating enzymes: a new class of biological regulators. *Crit Rev Biochem Mol Biol*, 337-352.

Davie, J.R. and Murphy, L.C. (1990) Level of ubiquitinated histone H2B in chromatin is coupled to ongoing transcription. *Biochemistry*, 4752-4757.

de The, H., Lavau, C., Marchio, A., Chomienne, C., Degos, L. and Dejean, A. (1991) The PML-RAR alpha fusion mRNA generated by the t(15;17)

translocation in acute promyelocytic leukemia encodes a functionally altered RAR. *Cell*, 675-684.

Dekel, N. (1996) Protein phosphorylation/dephosphorylation in the meiotic cell cycle of mammalian oocytes. *Rev Reprod*, 82-88.

Dennis, J.W., Granovsky, M. and Warren, C.E. (1999) Protein glycosylation in development and disease. *Bioessays*, 412-421.

Deshaies, R.J. (1999) SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu Rev Cell Dev Biol*, 435-467.

Desterro, J.M., Rodriguez, M.S. and Hay, R.T. (1998) SUMO-1 modification of IκBα inhibits NF-κB activation. *Mol Cell*, 233-239.

Desterro, J.M., Rodriguez, M.S., Kemp, G.D. and Hay, R.T. (1999) Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *J Biol Chem*, 10618-10624.

Desterro, J.M., Thomson, J. and Hay, R.T. (1997) Ubch9 conjugates SUMO but not ubiquitin. *FEBS Lett*, 297-300.

Duerksen-Hughes, P.J., Xu, X.X. and Wilkinson, K.D. (1987) Structure and function of ubiquitin: evidence for differential interactions of arginine-74

with the activating enzyme and the proteases of ATP-dependent proteolysis. *Biochemistry*, 6980-6987.

Dunphy, J.T., Greentree, W.K., Manahan, C.L. and Linder, M.E. (1996) G-protein palmitoyltransferase activity is enriched in plasma membranes. *J Biol Chem*, 7154-7159.

Duprez, E., Saurin, A.J., Desterro, J.M., Lallemand-Breitenbach, V., Howe, K., Boddy, M.N., Solomon, E., de The, H., Hay, R.T. and Freemont, P.S. (1999) SUMO-1 modification of the acute promyelocytic leukaemia protein PML: implications for nuclear localisation. *J Cell Sci*, 381-393.

Dutta, R., Qin, L. and Inouye, M. (1999) Histidine kinases: diversity of domain organization. *Mol Microbiol*, 633-640.

Dyck, J.A., Maul, G.G., Miller, W.H., Jr., Chen, J.D., Kakizuka, A. and Evans, R.M. (1994) A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. *Cell*, 333-343.

Edelman, A.M., Blumenthal, D.K. and Krebs, E.G. (1987) Protein serine/threonine kinases. *Annu Rev Biochem*, 567-613.

Fischer, E.H. (1999) Cell signaling by protein tyrosine phosphorylation. *Adv Enzyme Regul*, 359-369.

Fisk, H.A. and Yaffe, M.P. (1999) A role for ubiquitination in mitochondrial inheritance in Saccharomyces cerevisiae. *J Cell Biol*, 1199-1208.

Franz, J.K., Pap, T., Hummel, K.M., Nawrath, M., Aicher, W.K., Shigeyama, Y., Muller-Ladner, U., Gay, R.E. and Gay, S. (2000) Expression of sentrin, a novel antiapoptotic molecule, at sites of synovial invasion in rheumatoid arthritis. *Arthritis Rheum*, 599-607.

Furukawa, K., Mizushima, N., Noda, T. and Ohsumi, Y. (2000) A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J Biol Chem*, 7462-7465.

Galan, J. and Haguenauer-Tsapis, R. (1997) Ubiquitin lys63 is involved in ubiquitination of a yeast plasma membrane protein. *EMBO J*, 5847-5854.

Geyer, M., Herrmann, C., Wohlgemuth, S., Wittinghofer, A. and Kalbitzer, H.R. (1997) Structure of the Ras-binding domain of RalGEF and implications for Ras binding and signalling . *Nat Struct Biol*, 694-699.

Giorgino, F., de Robertis, O., Laviola, L., Montrone, C., Perrini, S., McCowen, K.C. and Smith, R.J. (2000) The sentrin-conjugating enzyme mUbc9 interacts with GLUT4 and GLUT1 glucose transporters and

regulates transporter levels in skeletal muscle cells. *Proc Natl Acad Sci U S A*, 1125-1130.

Giraud, M.F., Desterro, J.M. and Naismith, J.H. (1998) Structure of ubiquitin-conjugating enzyme 9 displays significant differences with other ubiquitin-conjugating enzymes which may reflect its specificity for sumo rather than ubiquitin. *Acta Crystallogr D Biol Crystallogr*, 891-898.

Gjertsen, B.T. and Doskeland, S.O. (1995) Protein phosphorylation in apoptosis. *Biochim Biophys Acta*, 187-199.

Goldberg, M.E., Expert-Bezancon, N., Vuillard, L. and Rabilloud, T. (1995) Non-detergent sulphobetaines: a new class of molecules that facilitate in vitro protein renaturation. *Fold Des*, 21-27.

Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D.H., Niall, H.D. and Boyse, E.A. (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc Natl Acad Sci U S A*, 11-15.

Gong, L., Kamitani, T., Fujise, K., Caskey, L.S. and Yeh, E.T. (1997) Preferential interaction of sentrin with a ubiquitin-conjugating enzyme, Ubc9. *J Biol Chem*, 28198-28201.

Gong, L., Kamitani, T., Millas, S. and Yeh, E.T. (2000a) Identification of a novel isopeptidase with dual specificity for ubiquitin- and NEDD8-conjugated proteins. *J Biol Chem*, 14212-14216.

Gong, L., Li, B., Millas, S. and Yeh, E.T. (1999) Molecular cloning and characterization of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. *FEBS Lett*, 185-189.

Gong, L., Millas, S., Maul, G.G. and Yeh, E.T. (2000b) Differential regulation of sentrinized proteins by a novel sentrin-specific protease. *J Biol Chem*, 3355-3359.

Gostissa, M., Hengstermann, A., Fogal, V., Sandy, P., Schwarz, S.E., Scheffner, M. and Del Sal, G. (1999) Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *EMBO J*, 6462-6471.

Gottlicher, M., Heck, S., Doucas, V., Wade, E., Kullmann, M., Cato, A.C., Evans, R.M. and Herrlich, P. (1996) Interaction of the Ubc9 human homologue with c-Jun and with the glucocorticoid receptor. *Steroids*, 257-262.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H.D. and Huber, R. (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature*, 463-471.

Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A.M. and Bax, A. (1992) 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry*, 8180-8190.

Gu, W. and Roeder, R.G. (1997) Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, 595-606.

Guex, N., Diemand, A., Schwede, T. and Peitsch, M.C. (1999) Swiss Pdb Viewer. Glaxo Wellcome Experimental Research.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 2714-2723.

Han, J.R., Suiko, M., Liu, C.C. and Liu, M.C. (1991) Post-translational modifications and binding properties of the apically secreted 80-kDa glycoprotein from Madin-Darby canine kidney cells: similarities to the C-terminal portion of the basolaterally secreted fibronectin. *Arch Biochem Biophys*, 337-345.

Hart, G.W. (1997) Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu Rev Biochem*, 315-335.

Hart, G.W., Haltiwanger, R.S., Holt, G.D. and Kelly, W.G. (1989) Nucleoplasmic and cytoplasmic glycoproteins. *Ciba Found Symp*, 102-112.

Hart, G.W., Holt, G.D. and Haltiwanger, R.S. (1988) Nuclear and cytoplasmic glycosylation: novel saccharide linkages in unexpected places. *Trends Biochem Sci*, 380-384.

Hay, R.T., Vuillard, L., Desterro, J.M. and Rodriguez, M.S. (1999) Control of NF-kappa B transcriptional activation by signal induced proteolysis of I kappa B alpha. *Philos Trans R Soc Lond B Biol Sci*, 1601-1609.

Heller, H. and Hershko, A. (1990) A ubiquitin-protein ligase specific for type III protein substrates. *J Biol Chem*, 6532-6535.

Hershko, A. and Ciechanover, A. (1998) The ubiquitin system. *Annu Rev Biochem*, 425-479.

Hiyama, H., Yokoi, M., Masutani, C., Sugasawa, K., Maekawa, T., Tanaka, K., Hoeijmakers, J.H. and Hanaoka, F. (1999) Interaction of hHR23 with S5a. The ubiquitin-like domain of hHR23 mediates interaction with S5a subunit of 26 S proteasome. *J Biol Chem*, 28019-28025.

Hochstrasser, M. (1996) Ubiquitin-dependent protein degradation. *Annu Rev Genet*, 405-439.

Hochstrasser, M. (2000) Evolution and function of ubiquitin-like protein-conjugation systems. *Nat Cell Biol*, E153-E157.

Hofmann, H., Floss, S. and Stamminger, T. (2000) Covalent modification of the transactivator protein IE2-p86 of human cytomegalovirus by conjugation to the ubiquitin-homologous proteins SUMO-1 and hSMT3b. *J Virol*, 2510-2524.

Holt, G.D., Snow, C.M., Senior, A., Haltiwanger, R.S., Gerace, L. and Hart, G.W. (1987) Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine. *J Cell Biol*, 1157-1164.

Huang, H.W., Tsoi, S.C., Sun, Y.H. and Li, S.S. (1998) Identification and characterization of the SMT3 cDNA and gene encoding ubiquitin-like protein from Drosophila melanogaster. *Biochem Mol Biol Int*, 775-785.

Huggins, G.S., Chin, M.T., Sibinga, N.E., Lee, S.L., Haber, E. and Lee, M.E. (1999) Characterization of the mUBC9-binding sites required for E2A protein degradation. *J Biol Chem*, 28690-28696.

Hunter, T. (1987) A thousand and one protein kinases. *Cell*, 823-829.

Hunter, T. and Cooper, J.A. (1985) Protein-tyrosine kinases. *Annu Rev Biochem*, 897-930.

Imperiali, B. and O'Connor, S.E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol*, 643-649.

Ishov, A.M., Sotnikov, A.G., Negorev, D., Vladimirova, O.V., Neff, N., Kamitani, T., Yeh, E.T., Strauss, J.F., 3rd and Maul, G.G. (1999) PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. *J Cell Biol*, 221-234.

Jaffray, E., Wood, K.M. and Hay, R.T. (1995) Domain organization of I kappa B alpha and sites of interaction with NF-kappa B p65. *Mol Cell Biol*, 2166-2172.

Jentsch, S. (1992) The ubiquitin-conjugation system. *Annu Rev Genet*, 179-207.

Jentsch, S. and Pyrowolakis, G. (2000) Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol*, 335-342.

Johnson, E.S. and Blobel, G. (1997) Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J Biol Chem*, 26799-26802.

Johnson, E.S. and Blobel, G. (1999) Cell cycle-regulated attachment of the ubiquitin-related protein SUMO to the yeast septins. *J Cell Biol*, 981-994.

Johnson, E.S., Schwienhorst, I., Dohmen, R.J. and Blobel, G. (1997) The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *EMBO J*, 5509-5519.

Johnston, S.C., Larsen, C.N., Cook, W.J., Wilkinson, K.D. and Hill, C.P. (1997) Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 A resolution. *EMBO J*, 3787-3796.

Kalhammer, G., Bahler, M., Schmitz, F., Jockel, J. and Block, C. (1997) Ras-binding domains: predicting function versus folding. *FEBS Lett*, 599-602.

Kamitani, T., Kito, K., Nguyen, H.P., Fukuda-Kamitani, T. and Yeh, E.T. (1998a) Characterization of a second member of the sentrin family of ubiquitin-like proteins. *J Biol Chem*, 11349-11353.

Kamitani, T., Kito, K., Nguyen, H.P., Wada, H., Fukuda-Kamitani, T. and Yeh, E.T. (1998b) Identification of three major sentrinization sites in PML. *J Biol Chem*, 26675-26682.

Kamitani, T., Kito, K., Nguyen, H.P. and Yeh, E.T. (1997a) Characterization of NEDD8, a developmentally down-regulated ubiquitin-like protein. *J Biol Chem*, 28557-28562.

Kamitani, T., Nguyen, H.P., Kito, K., Fukuda-Kamitani, T. and Yeh, E.T. (1998c) Covalent modification of PML by the sentrin family of ubiquitin-like proteins. *J Biol Chem*, 3117-3120.

Kamitani, T., Nguyen, H.P. and Yeh, E.T. (1997b) Preferential modification of nuclear proteins by a novel ubiquitin-like molecule. *J Biol Chem*, 14001-14004.

Kamura, T., Conrad, M.N., Yan, Q., Conaway, R.C. and Conaway, J.W. (1999) The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2. *Genes Dev*, 2928-2933.

Kasinathan, C., Grzelinska, E., Okazaki, K., Slomiany, B.L. and Slomiany, A. (1990) Purification of protein fatty acyltransferase and determination of its distribution and topology. *J Biol Chem*, 5139-5144.

Kawabe, Y., Seki, M., Seki, T., Wang, W.S., Imamura, O., Furuichi, Y., Saitoh, H. and Enomoto, T. (2000) Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1. *J Biol Chem*, 20963-20966.

Kearse, K.P. and Hart, G.W. (1991) Topology of O-linked N-acetylglucosamine in murine lymphocytes. *Arch Biochem Biophys*, 543-548.

Kemp, B.E. and Pearson, R.B. (1990) Protein kinase recognition sequence motifs. *Trends Biochem Sci*, 342-346.

Kim, K.I., Baek, S.H., Jeon, Y.J., Nishimori, S., Suzuki, T., Uchida, S., Shimbara, N., Saitoh, H., Tanaka, K. and Chung, C.H. (2000) A New SUMO-1-specific Protease, SUSP1, That Is Highly Expressed in Reproductive Organs. *J Biol Chem*, 14102-14106.

Kim, Y.H., Choi, C.Y. and Kim, Y. (1999) Covalent modification of the homeodomain-interacting protein kinase 2 (HIPK2) by the ubiquitin-like protein SUMO-1. *Proc Natl Acad Sci U S A*, 12350-12355.

Klionsky, D.J. and Ohsumi, Y. (1999) Vacuolar import of proteins and organelles from the cytoplasm. *Annu Rev Cell Dev Biol*, 1-32.

Ko, L.J. and Prives, C. (1996) p53: puzzle and paradigm. *Genes Dev*, 1054-1072.

Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H.D., Mayer, T.U. and Jentsch, S. (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell*, 635-644.

Kornfield, S. (1998) Diseases of abnormal protein glycosylation: an emerging area. *J Clin Invest*, 1293-1295.

Kouzarides, T. (2000) Acetylation: a regulatory modification to rival phosphorylation? *EMBO J*, 1176-1179.

Krebs, E.G., Graves, D.J. and Fischer, E.H. (1959) Factors affecting the activity of muscle phosphorylase b-kinase. *J. Biol. Chem.*, 2867-2873.

Krishna, R.G. and Wold, F. (1993) Post-translational modification of proteins. *Adv Enzymol Relat Areas Mol Biol*, 265-298.

Lammer, D., Mathias, N., Laplaza, J.M., Jiang, W., Liu, Y., Callis, J., Goebl, M. and Estelle, M. (1998) Modification of yeast Cdc53p by the ubiquitin-related protein rub1p affects function of the SCFCdc4 complex. *Genes Dev*, 914-926.

Lapenta, V., Chiurazzi, P., van der Spek, P., Pizzuti, A., Hanaoka, F. and Brahe, C. (1997) SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to chromosome 21qter and defines a novel gene family. *Genomics*, 362-366.

Larsen, C.N., Price, J.S. and Wilkinson, K.D. (1996) Substrate binding and catalysis by ubiquitin C-terminal hydrolases: identification of two active site residues. *Biochemistry*, 6735-6744.

Lau, K.H., Farley, J.R. and Baylink, D.J. (1989) Phosphotyrosyl protein phosphatases. *Biochem J,* 23-36.

Lehembre, F., Badenhorst, P., Muller, S., Travers, A., Schweisguth, F. and Dejean, A. (2000) Covalent modification of the transcriptional repressor tramtrack by the ubiquitin-related protein Smt3 in Drosophila flies. *Mol Cell Biol,* 1072-1082.

Levitan, I.B. (1999) Modulation of ion channels by protein phosphorylation. How the brain works. *Adv Second Messenger Phosphoprotein Res,* 3-22.

Li, S.J. and Hochstrasser, M. (1999) A new protease required for cell-cycle progression in yeast. *Nature,* 246-251.

Li, S.J. and Hochstrasser, M. (2000) The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Mol Cell Biol,* 2367-2377.

Liakopoulos, D., Busgen, T., Brychzy, A., Jentsch, S. and Pause, A. (1999) Conjugation of the ubiquitin-like protein NEDD8 to cullin-2 is linked to von Hippel-Lindau tumor suppressor function. *Proc Natl Acad Sci U S A,* **96**, 5510-5515.

Liakopoulos, D., Doenges, G., Matuschewski, K. and Jentsch, S. (1998) A novel protein modification pathway related to the ubiquitin system. *EMBO J*, 2208-2214.

Lin, H. and Wing, S.S. (1999) Identification of rabbit reticulocyte E217K as a UBC7 homologue and functional characterization of its core domain loop. *J Biol Chem*, 14685-14691.

Liu, J.P. (1997) Protein phosphorylation events in exocytosis and endocytosis. *Clin Exp Pharmacol Physiol*, 611-618.

Liu, Y.C., Pan, J., Zhang, C., Fan, W., Collinge, M., Bender, J.R. and Weissman, S.M. (1999) A MHC-encoded ubiquitin-like protein (FAT10) binds noncovalently to the spindle assembly checkpoint protein MAD2. *Proc Natl Acad Sci U S A*, 4313-4318.

Loeb, K.R. and Haas, A.L. (1994) Conjugates of ubiquitin cross-reactive protein distribute in a cytoskeletal pattern. *Mol Cell Biol*, 8408-8419.

Long, X. and Griffith, L.C. (2000) Identification and Characterization of a SUMO-1 Conjugation System That Modifies Neuronal CaMKII in Drosophila melanogaster. *J Biol Chem*.

Magnaghi-Jaulin, L., Ait-Si-Ali, S. and Harel-Bellan, A. (1999) Histone acetylation in signal transduction by growth regulatory signals. *Semin Cell Dev Biol*, 197-203.

Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell*, 97-107.

Mahajan, R., Gerace, L. and Melchior, F. (1998) Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. *J Cell Biol*, 259-270.

Maki, C.G., Huibregtse, J.M. and Howley, P.M. (1996) In vivo ubiquitination and proteasome-mediated degradation of p53(1). *Cancer Res*, 2649-2654.

Mannen, H., Tseng, H.M., Cho, C.L. and Li, S.S. (1996) Cloning and expression of human homolog HSMT3 to yeast SMT3 suppressor of MIF2 mutations in a centromere protein gene. *Biochem Biophys Res Commun*, 178-180.

Mao, Y., Desai, S.D. and Liu, L.F. (2000a) SUMO-1 conjugation to human DNA topoisomerase II isozymes. *J Biol Chem*, 26066-26073.

Mao, Y., Sun, M., Desai, S.D. and Liu, L.F. (2000b) SUMO-1 conjugation to topoisomerase I: A possible repair response to topoisomerase-mediated DNA damage. *Proc Natl Acad Sci U S A*, 4046-4051.

Martinez-Noel, G., Niedenthal, R., Tamura, T. and Harbers, K. (1999) A family of structurally related RING finger proteins interacts specifically with the ubiquitin-conjugating enzyme UbcM4. *FEBS Lett*, 257-261.

Matunis, M.J., Coutavas, E. and Blobel, G. (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*, 1457-1470.

Matunis, M.J., Wu, J. and Blobel, G. (1998) SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *J Cell Biol*, 499-509.

Meek, D.W., Campbell, L.E., Jardine, L.J., Knippschild, U., McKendrick, L. and Milne, D.M. (1997) Multi-site phosphorylation of p53 by protein kinases inducible by p53 and DNA damage. *Biochem Soc Trans*, 416-419.

Melchior, F., Paschal, B., Evans, J. and Gerace, L. (1993) Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. *J Cell Biol*, 1649-1659.

Meluh, P.B. and Koshland, D. (1995) Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Mol Biol Cell*, 793-807.

Minty, A., Dumont, X., Kaghad, M. and Caput, D. (2000) Covalent Modification of p73alpha by SUMO-1. TWO-HYBRID SCREENING WITH p73 IDENTIFIES NOVEL SUMO-1-INTERACTING PROTEINS AND A SUMO-1 INTERACTION MOTIF. *J Biol Chem*, 36316-36323.

Miura, T., Klaus, W., Gsell, B., Miyamoto, C. and Senn, H. (1999) Characterisation of the binding interface between ubiquitin and class I human ubiquitin-conjugating enzyme 2b by multidimensional heteronuclear NMR spectroscopy in solution. *J. Mol. Biol.* 213-228.

Moore, M.S. and Blobel, G. (1993) The GTP-binding protein Ran/TC4 is required for protein import into the nucleus. *Nature*, 661-663.

Morett, E. and Bork, P. (1999) A novel transactivation domain in parkin. *Trends Biochem Sci*, 229-231.

Mossessova, E. and Lima, C.D. (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell*, 865-876.

Moudgil, V.K. (1989) Receptors Phosphorylation. *CRC Press, Fla.*

Muller, S., Berger, M., Lehembre, F., Seeler, J.S., Haupt, Y. and Dejean, A. (2000) c-Jun and p53 Activity Is Modulated by SUMO-1 Modification. *J Biol Chem*, 13321-13329.

Muller, S. and Dejean, A. (1999) Viral immediate-early proteins abrogate the modification by SUMO-1 of PML and Sp100 proteins, correlating with nuclear body disruption. *J Virol*, 5137-5143.

Muller, S., Matunis, M.J. and Dejean, A. (1998a) Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO J*, 61-70.

Muller, S., Miller, W.H., Jr. and Dejean, A. (1998b) Trivalent antimonials induce degradation of the PML-RAR oncoprotein and reorganization of the promyelocytic leukemia nuclear bodies in acute promyelocytic leukemia NB4 cells. *Blood*, 4308-4316.

Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCormick, F. and Wittinghofer, A. (1995) The 2.2 A crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature*, 554-560.

Nigg, E.A. (1997) Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature*, 779-787.

Okuma, T., Honda, R., Ichikawa, G., Tsumagari, N. and Yasuda, H. (1999) In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. *Biochem Biophys Res Commun*, 693-698.

Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C.F., Chang, H.M. and Yeh, E.T. (1996) Protection against Fas/APO-1- and tumor necrosis factor-mediated cell death by a novel protein, sentrin. *J Immunol*, 4277-4281.

Orengo, C.A., Jones, D.T. and Thornton, J.M. (1994) Protein superfamilies and domain superfolds. *Nature*, 631-634.

Osaka, F., Kawasaki, H., Aida, N., Saeki, M., Chiba, T., Kawashima, S., Tanaka, K. and Kato, S. (1998) A new NEDD8-ligating system for cullin-4A. *Genes Dev*, 2263-2268.

Patarca, R. (1996) Protein phosphorylation and dephosphorylation in physiologic and oncologic processes. *Crit Rev Oncog*, 343-432.

Peitsch, M.C., Herzyk, P., Wells, T.N. and Hubbard, R.E. (1996) Automated modelling of the transmembrane region of G-protein coupled receptor by Swiss-model. *Receptors Channels*, 161-164.

Pickart, C.M. (2000) Ubiquitin in chains. *Trends Biochem Sci*, 544-548.

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR*, 661-665.

Pogo, B.G., Allfrey, V.G. and Mirsky, A.E. (1966) RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proc Natl Acad Sci U S A*, 805-812.

Poukka, H., Aarnisalo, P., Karvonen, U., Palvimo, J.J. and Janne, O.A. (1999) Ubc9 interacts with the androgen receptor and activates receptor-dependent transcription. *J Biol Chem*, 19441-19446.

Rajagopalan, K.V. (1997) Biosynthesis and processing of the molybdenum cofactors. *Biochem Soc Trans*, 757-761.

Rangasamy, D. and Wilson, V.G. (2000) Bovine papillomavirus E1 protein is sumoylated by the host cell ubc9 protein. *J Biol Chem*, 30487-30495.

Rangasamy, D., Woytek, K., Khan, S.A. and Wilson, V.G. (2000) SUMO-1 modification of bovine papillomavirus E1 protein is required for intranuclear accumulation. *J Biol Chem*, 37999-38004.

Rodriguez, M.S., Desterro, J.M., Lain, S., Midgley, C.A., Lane, D.P. and Hay, R.T. (1999) SUMO-1 modification activates the transcriptional response of p53. *EMBO J*, 6455-6461.

Saitoh, H. and Hinchey, J. (2000) Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem*, 6252-6258.

Saitoh, H., Sparrow, D.B., Shiomi, T., Pu, R.T., Nishimoto, T., Mohun, T.J. and Dasso, M. (1998) Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2. *Curr Biol*, 121-124.

Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W. and Madura, K. (1998) Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature*, 715-718.

Schneider, B.L., Yang, Q.H. and Futcher, A.B. (1996) Linkage of replication to start by the Cdk inhibitor Sic1. *Science*, 560-562.

Schwartz, A.L. and Ciechanover, A. (1999) The ubiquitin-proteasome pathway and pathogenesis of human diseases. *Annu Rev Med*, 57-74.

Schwarz, S.E., Matuschewski, K., Liakopoulos, D., Scheffner, M. and Jentsch, S. (1998) The ubiquitin-like proteins SMT3 and SUMO-1 are conjugated by the UBC9 E2 enzyme. *Proc Natl Acad Sci U S A*, 560-564.

Schwienhorst, I., Johnson, E.S. and Dohmen, R.J. (2000) SUMO conjugation and deconjugation. *Mol Gen Genet*, 771-786.

Seeler, J.S. and Dejean, A. (1999) The PML nuclear bodies: actors or extras? *Curr Opin Genet Dev*, 362-367.

Shao, W., Fanelli, M., Ferrara, F.F., Riccioni, R., Rosenauer, A., Davison, K., Lamph, W.W., Waxman, S., Pelicci, P.G., Lo Coco, F., Avvisati, G., Testa, U., Peschle, C., Gambacorti-Passerini, C., Nervi, C. and Miller, W.H., Jr. (1998) Arsenic trioxide as an inducer of apoptosis and loss of PML/RAR alpha protein in acute promyelocytic leukemia cells. *J Natl Cancer Inst*, 124-133.

Shaw, P., Freeman, J., Bovey, R. and Iggo, R. (1996) Regulation of specific DNA binding by p53: evidence for a role for O-glycosylation and charged residues at the carboxy terminus. *Oncogene*, 921-930.

Shen, Z., Pardington-Purtymun, P.E., Comeaux, J.C., Moyzis, R.K. and Chen, D.J. (1996) UBL1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins. *Genomics*, 271-279.

Silver, E.T., Gwozd, T.J., Ptak, C., Goebl, M. and Ellison, M.J. (1992) A chimeric ubiquitin conjugating enzyme that combines the cell cycle properties of CDC34 (UBC3) and the DNA repair properties of RAD6 (UBC2): implications for the structure, function and evolution of the E2s. *EMBO J*, 3091-3098.

Sklenar, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in 15N,13C-labeled RNA oligonucleotides. *J Biomol NMR*, 721-727.

Snow, D.M. and Hart, G.W. (1998) Nuclear and cytoplasmic glycosylation. *Int Rev Cytol*, 43-74.

Sommer, T. and Wolf, D.H. (1997) Endoplasmic reticulum degradation: reverse protein flow of no return. *Faseb J*, 1227-1233.

Soyombo, A.A. and Hofmann, S.L. (1997) Molecular cloning and expression of palmitoyl-protein thioesterase 2 (PPT2), a homolog of lysosomal palmitoyl-protein thioesterase with a distinct substrate specificity. *J Biol Chem*, 27456-27463.

Spence, J., Gali, R.R., Dittmar, G., Sherman, F., Karin, M. and Finley, D. (2000) Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell*, 67-76.

Spence, J., Sadis, S., Haas, A.L. and Finley, D. (1995) A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol*, 1265-1273.

Spencer, V.A. and Davie, J.R. (1999) Role of covalent modifications of histones in regulating gene expression. *Gene*, 1-12.

Sterner, D.E. and Berger, S.L. (2000) Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev*, 435-459.

Sternsdorf, T., Jensen, K., Reich, B. and Will, H. (1999) The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. *J Biol Chem*, 12555-12566.

Sternsdorf, T., Jensen, K. and Will, H. (1997) Evidence for covalent modification of the nuclear dot-associated proteins PML and Sp100 by PIC1/SUMO-1. *J Cell Biol*, 1621-1634.

Stochaj, U., Hejazi, M. and Belhumeur, P. (1998) The small GTPase Gsp1p binds to the repeat domain of the nucleoporin Nsp1p. *Biochem J*, 421-427.

Struhl, K. (1998) Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev*, **12**, 599-606.

Stryer, L. (1995) *Biochemistry*. W. H. Freeman & Co., New York.

Stuurman, N., de Graaf, A., Floore, A., Josso, A., Humbel, B., de Jong, L. and van Driel, R. (1992) A monoclonal antibody recognizing nuclear matrix-associated nuclear bodies. *J Cell Sci*, 773-784.

Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F.C., Ruderman, J.V. and Hershko, A. (1995) The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol Biol Cell*, 185-197.

Suopanki, J., Tyynela, J., Baumann, M. and Haltia, M. (1999) The expression of palmitoyl-protein thioesterase is developmentally regulated in neural tissues but not in nonneural tissues. *Mol Genet Metab*, 290-293.

Suzuki, T., Ichiyama, A., Saitoh, H., Kawakami, T., Omata, M., Chung, C.H., Kimura, M., Shimbara, N. and Tanaka, K. (1999) A new 30-kDa ubiquitin-related SUMO-1 hydrolase from bovine brain. *J Biol Chem*, 31131-31134.

Swope, S.L., Moss, S.I., Raymond, L.A. and Huganir, R.L. (1999) Regulation of ligand-gated ion channels by protein phosphorylation. *Adv Second Messenger Phosphoprotein Res*, **33**, 49-78.

Takahashi, Y., Iwase, M., Konishi, M., Tanaka, M., Toh-e, A. and Kikuchi, Y. (1999) Smt3, a SUMO-1 homolog, is conjugated to Cdc3, a component of septin rings at the mother-bud neck in budding yeast. *Biochem Biophys Res Commun*, 582-587.

Takahashi, Y., Mizoi, J., Toh, E.A. and Kikuchi, Y. (2000) Yeast Ulp1, an Smt3-Specific Protease, Associates with Nucleoporins. *J Biochem (Tokyo)*, 723-725.

Tanaka, K., Nishide, J., Okazaki, K., Kato, H., Niwa, O., Nakagawa, T., Matsuda, H., Kawamukai, M. and Murakami, Y. (1999) Characterization of a fission yeast SUMO-1 homologue, pmt3p, required for multiple nuclear events, including the control of telomere length and chromosome segregation. *Mol Cell Biol*, 8660-8672.

Tanaka, K., Suzuki, T. and Chiba, T. (1998) The ligation systems for ubiquitin and ubiquitin-like proteins. *Mol Cells*, 503-512.

Taylor, S.V., Kelleher, N.L., Kinsland, C., Chiu, H.J., Costello, C.A., Backstrom, A.D., McLafferty, F.W. and Begley, T.P. (1998) Thiamin biosynthesis in Escherichia coli. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J Biol Chem*, 16555-16560.

Terrell, J., Shih, S., Dunn, R. and Hicke, L. (1998) A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol Cell*, 193-202.

Tong, H., Hateboer, G., Perrakis, A., Bernards, R. and Sixma, T.K. (1997) Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. *J Biol Chem*, 21381-21387.

Torres, C.R. and Hart, G.W. (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem*, 3308-3317.

Tyers, M. and Willems, A.R. (1999) One ring to rule a superfamily of E3 ubiquitin ligases. *Science*, 601, 603-604.

Unkles, S.E., Heck, I.S., Appleyard, M.V. and Kinghorn, J.R. (1999) Eukaryotic molybdopterin synthase. Biochemical and molecular studies of Aspergillus nidulans cnxG and cnxH mutants. *J Biol Chem*, 19286-19293.

Verkruyse, L.A. and Hofmann, S.L. (1996) Lysosomal targeting of palmitoyl-protein thioesterase. *J Biol Chem*, 15831-15836.

Vesa, J., Hellsten, E., Verkruyse, L.A., Camp, L.A., Rapola, J., Santavuori, P., Hofmann, S.L. and Peltonen, L. (1995) Mutations in the palmitoyl

protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature*, 584-587.

Wada, H., Kito, K., Caskey, L.S., Yeh, E.T.H. and Kamitani, T. (1998) Cleavage of the C-terminus of NEDD8 by UCH-L3. *Biochem Biophys Res Commun*, 688-692.

Wada, H., Yeh, E.T. and Kamitani, T. (1999) The von Hippel-Lindau tumor suppressor gene product promotes, but is not essential for, NEDD8 conjugation to cullin-2. *J Biol Chem*, 36025-36029.

Watkins, J.F., Sung, P., Prakash, L. and Prakash, S. (1993) The Saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol Cell Biol*, 7757-7765.

Wedegaertner, P.B., Wilson, P.T. and Bourne, H.R. (1995) Lipid modifications of trimeric G proteins. *J Biol Chem*, 503-506.

Wellner, D., Panneerselvam, C. and Horecker, B.L. (1990) Sequencing of peptides and proteins with blocked N-terminal amino acids: N-acetylserine or N-acetylthreonine. *Proc Natl Acad Sci U S A*, 1947-1949.

Whitby, F.G., Xia, G., Pickart, C.M. and Hill, C.P. (1998) Crystal structure of the human ubiquitin-like protein NEDD8 and interactions with ubiquitin pathway enzymes. *J Biol Chem*, 34983-34991.

Wilkinson, K.D. (1997a) Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *Faseb J*, 1245-1256.

Wilkinson, K.D. (1997b) *Ubiquitin Carboxyl-terminal Hydrolase. A Handbook of Proteolytic Enzymes*. Academic Press, London.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 1647-1651.

Wolffe, A.P. (1992) New insights into chromatin function in transcriptional control. *Faseb J*, 3354-3361.

Xu, W., Gong, L., Haddad, M.M., Bischof, O., Campisi, J., Yeh, E.T. and Medrano, E.E. (2000) Regulation of microphthalmia-associated transcription factor MITF protein levels by association with the ubiquitin-conjugating enzyme hUBC9. *Exp Cell Res*, 135-143.

Yang, M., Ellenberg, J., Bonifacino, J.S. and Weissman, A.M. (1997) The transmembrane domain of a carboxyl-terminal anchored protein

determines localization to the endoplasmic reticulum. *J Biol Chem*, 1970-1975.

Yeh, E.T., Gong, L. and Kamitani, T. (2000) Ubiquitin-like proteins: new wines in new bottles. *Gene*, 1-14.

Zheng, N., Wang, P., Jeffrey, P.D. and Pavletich, N.P. (2000) Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell*, 533-539.

7. PUBLICATIONS

Tatham, M. H., Naismith, J. H. and Hay, R. T. Role of SUMO-1 C-terminus in the enzymatic linkage to substrate. Under submission to J. Biol. C.

Tatham, M. H., Desterro, M. J. P., Naismith, J. H. and Hay, R. T. SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. Under submission to FEBBS Letts.