Interactions of SUMO proteins

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

Dr. rer. nat.

des Fachbereichs
Bio- und Geowissenschaften,
Landschaftsarchitektur

an der

Universität Duisburg-Essen

vorgelegt von

Matthias Rabiller

aus Den Haag

Oktober 2005

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden am Max Planck Institut für molekulare Physiologie (Dortmund) und in der Abteilung für strukturelle und medizinische Biochemie der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Peter Bayer

2. Gutachter: Prof. Dr. Tarik Moröy

Vorsitzender des Prüfungsausschusses: Prof. Dr. Helmut Esche

Tag der mündlichen Prüfung: 16. Juni 2006

Content

Abstract

A large number of proteins have shown ability to bind to SUMO (Small Ubiquitin like Modifier) proteins through a short conserved motif called SIM (SUMO Interacting Motif).

The work presented here shows that the different SUMO isoforms interact with the hydrophobic core of the SIM by forming an intermolecular β-sheet with the β2 strand of SUMO. This interaction is crucial for SUMO binding, and is modulated by interactions between SUMO and the amino acids flanking the core of the SIM. The SIM can be phosphorylated, providing a possibility for regulating the strength of SUMO binding in the lifetime of a protein. Furthermore, a concentration threshold effect is observed in the binding of the unphosphorylated SIM of PIAS (Protein inhibitor of activated STAT) to SUMO. The dependency on the amino acids flanking the hydrophobic core is stronger in binding to SUMO1 than to SUMO2, providing a mechanism for SUMO isoform discrimination.

Zusammenfassung

Viele Proteine haben die Fähigkeit an SUMO (Small Ubiquitin like Modifier) Proteine durch ein kleines konserviertes Motiv zu binden. Dieses Motiv wird SIM (SUMO Interacting Motif) genannt.

Diese Arbeit zeigt, dass der hydrophobe Kern des SIM mit den verschiedenen SUMO Isoformen interagiert, indem er mit dem β2 Strang von SUMO ein intermolekulares β-Faltblatt formt. Diese Wechselwirkung ist für die Bindung an SUMO essenziell und wird von Interaktionen zwischen SUMO und Aminosäuren, welche das SIM flankieren, verstärkt. Das SIM kann phosphoryliert werden, wodurch die Affinität der Bindung an SUMO während der Lebenszeit eines Proteins reguliert werden kann. Die flankierenden Aminosäuren des SIMs spielen eine größere Rolle bei der Bindung an SUMO1 als an SUMO2, was einen SUMO Isoform Diskriminierungsmechanismus darstellt.

Abbreviations

cDNA: Complementary Desoxyribonucleic Acid
EST: Expressed Sequence Tag
FID: Free Induced Decay
GSH: Glutathion (reduced form)
HEK293T: Human Embryonic Kidney cell line in which the gene for the temperature sensitive SV-40 T antigene was inserted.
HSQC: Heteronuclear Single Quantum Coherence
IPTG: Isopropyl-β-D-Thiogalactopyranoside
K_e: exhange constant (equivalent to K_D for individual atoms or chemical groups in a macromolecule)
K_D: Dissociation constant
Kon: association rate
Kex: dissociation rate
LB: Luria Bertani
MALDI: Matrix Assisted Laser Desorbtion Ionisation
NMR: Nuclear Magnetic Resonance
PBS: Phosphate Buffer Saline
PCR: Polymerase Chain Reaction
PDB: Protein DataBase (see www.rcsb.org)
PIAS: Protein Inhibitor of Activated STAT
SDS-PAGE: Sodium Dodecyl Sulfate – Poly Acrylamide Gel Electrophoresis
SIM: SUMO Interacting Motif
SUMO: Small Ubiquitin like Modifier
TDG: Thymin DNA Glycosidase
TTRAP: TRAF and TNF Receptor Associated Protein
2YT: Yeast Trypton, 2x concentrated

Introduction

I) Nuclear Magnetic Resonance

NMR (Nuclear Magnetic Resonance) is the primary method used in the work presented here. It will therefore be presented shortly in this section. NMR is to date the only available technique to observe the behavior of molecules in solution at atomic resolution. It relies on a property of subatomic particles called Spin. Spin is a property of particles, like mass, charge or magnetism. But unlike mass, charge or magnetism, Spin is not observable in the macroscopic world, making it impossible for us to have a "real life" experience of it. This makes Spin physics especially difficult to understand. Too many exciting things happen when a molecule is placed in a pulsed magnetic field to be described in detail in the introduction of a PhD thesis. The following section will describe only the basic principles of the experiments used in the present work.

1. Nature of Spin

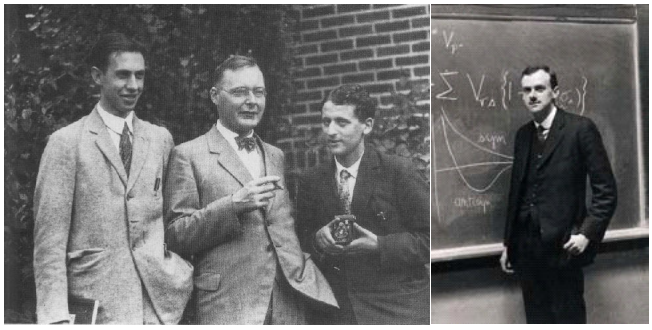


Fig. 1: from left to right George Uhlenbeck, Hendrik Kramers, Samuel Goudsmit and Paul Dirac

The existence of Spin was proposed in 1925 by Goudsmit and Uhlenbeck, and demonstrated in 1928 by Dirac (fig. 1). They showed that electrons behave *as if* they were tiny charged balls rotating about their own axis and thus creating, due to their charge, a magnetic field. This behavior is the manifestation of the Spin of the electrons, and has nothing to do with actual rotation of the electrons around themselves – even if this image is widely used in textbooks. Like electrons, neutrons and protons also have Spin.

Spin is quantified, comes in multiples of ½ and can be either positive or negative. This means the value of the Spin of a particle is either +½ or -½. In atoms, electrons, protons and neutrons fill orbitals following the "Aufbau principle". As a result, atoms have a net electronic Spin –the sum of the unpaired electron Spins- and a net nuclear Spin –the sum of the unpaired protons and neutrons Spins. Some

Nuclei have no Spin, others have a Spin ½, and others have even higher Spin (1, 3/2, etc…).

2. Principle of NMR

Placed in a magnetic field, a nucleus with Spin ½ can occupy two different energy states: Spin up (+½) and Spin down (-½). A convenient visualization is to consider such an nucleus as a magnet placed parallel (Spin down) or antiparallel (Spin up) to an exterior magnetic field. The energy difference between both states depends on the strength of the exterior magnetic field (fig. 2), and on the type of Nucleus (^{13}C, ^{1}H, ^{15}N…) Transition between those states is spontaneous. When a nucleus is placed in a magnetic field, its Spin tends to point in the same direction as the surrounding magnetic field, and transition to the opposite direction is possible upon irradiation of the nucleus with a radiofrequency which wavelength corresponds to the difference of energy between the two Spin states. An accurate description of the behavior of individual Spins in the course of NMR experiments is out of the scope of this introduction. What matters in an experiment is the behaviour of a population of Spins as a whole. Placed in an exterior magnetic field B_0, a population of Spins produces a magnetic field B_1 parallel to B_0. By using an appropriate sequence of radiofrequency excitations ("pulse sequence), B_1 can be tipped at 90° relatively to B_0. After the end of the pulse sequence, B_1 returns to its original position in a complex motion that involves turning about B_0 and bringing the (B_0, B_1) angle from 90° to 0° (fig. 3). This creates an alternative current of decreasing intensity in a coil placed perpendicularly to B_0, which is the signal measured in NMR. This signal is called FID (Free Induced decay). Fourier transformation can be used to

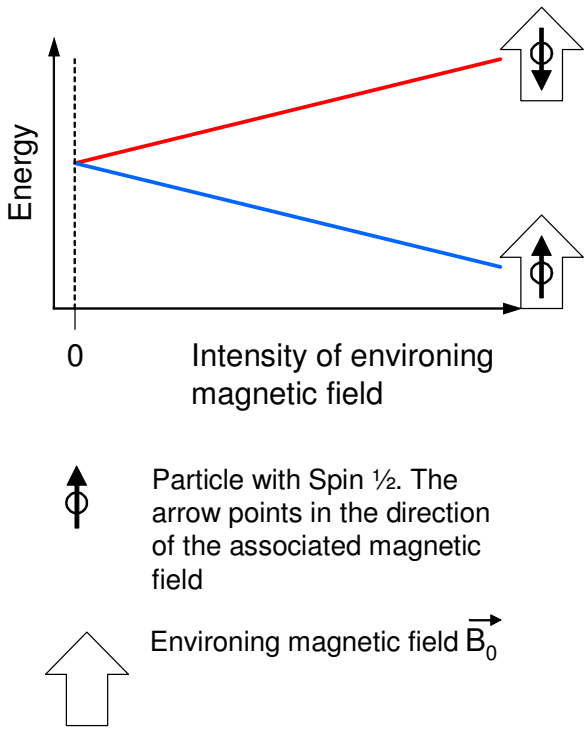


Fig. 2: the energy difference between the two states of a Spin –parallel to the environing magnetic field (blue) or antiparallel to it (red)- depends on the strength of this field.

extract the frequencies contained in the FID and produced a NMR spectrum in which

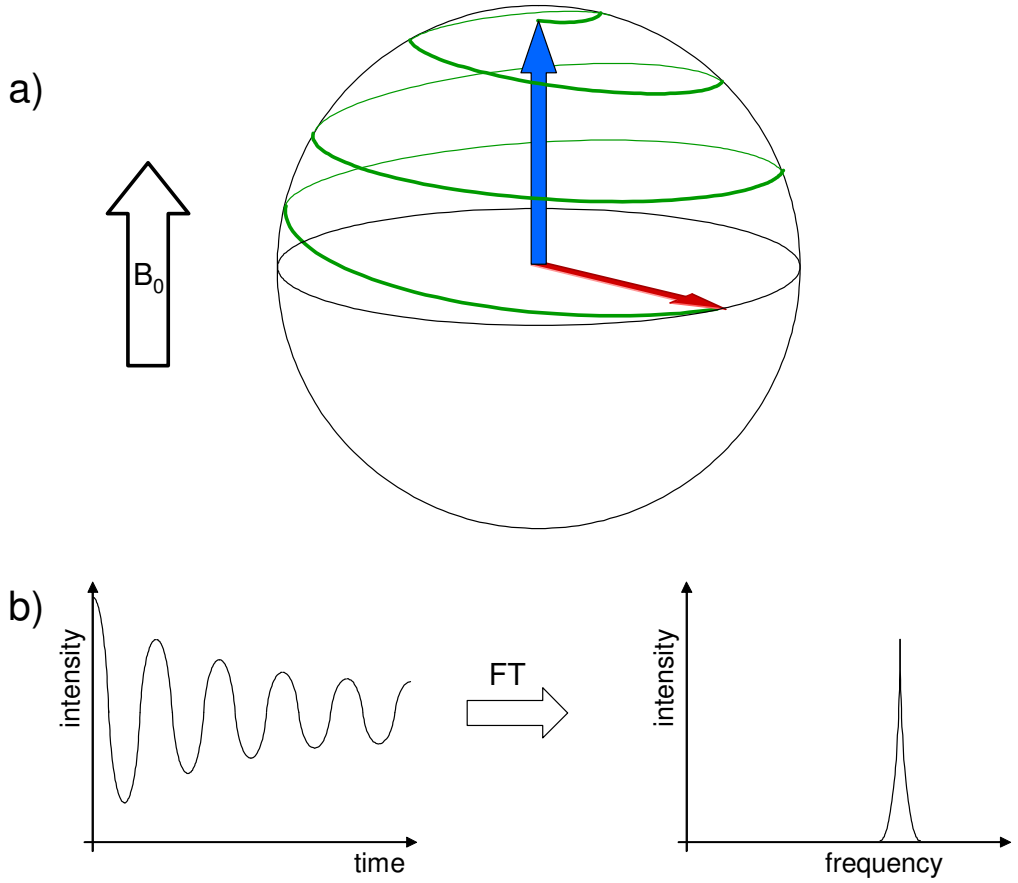


Fig. 3: a) After a pulse, the a population of Spin that has been oriented perpendicularly to the surrounding magnetic field B_0 (red arrow) comes back to its equilibrium position parallel to B_0 (blue arrow), following the path shown in green. This movement of the Spin population induces an oscillating signal (b, left) called FID in a the detector coil placed perpendicularly to B_0. Fourier transformation (FT) is used to convert the FID into a NMR spectrum (b, right).

the peaks correspond to the absorbed by the Spin populations present in the sample (resonance frequencies). Expressed in Hz, those frequencies are very large, and depend on the strength of the magnetic field emitted by the spectrometer. This is why one preferably expresses the resonance frequencies in ppm.

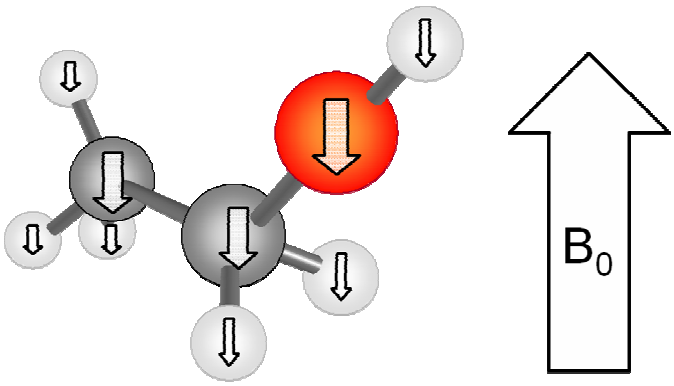


Fig. 4: A magnetic field B_0 induces the electronic clouds of the different atoms of an ethanol molecule to create magnetic fields of opposite direction. Each nucleus experiences the local sum of the applied field and the various induced fields.

3. 1D NMR

If the resonance of nuclei was dependent only on the type of nuclei and the magnetic field applied by the spectrometer, all

the hydrogen atoms of a macromolecule such as a protein would give the same signal, and no information on the macromolecule could be gained. This is, however, not the case. Since the electrons of the molecule are able to move, the applied magnetic field causes them to do so in a way that creates a magnetic field that counters the applied one, as described by the induction law. The intensity of this induced magnetic field varies within the molecule. Thus, each nucleus in a molecule is subjected to a magnetic field created by the atoms surrounding it (fig. 4). During an NMR experiment, a nucleus doesn't "feel" only the magnetic field created by the spectrometer, but the sum of the local magnetic field created by the molecule and the magnetic field generated by the spectrometer. Therefore, the different ^{13}C, ^{1}H and ^{15}N nuclei of a molecule absorb and reemit photons of different energies (they have different resonance frequencies) giving distinct NMR signals (fig. 5). In consequence, the ^{1}H NMR spectrum of a protein contains as many peaks as there are hydrogen atoms in that protein. There are, of course, many of them, and such a spectrum is very complex, with most peaks overlapping each other. To overcome this complexity,

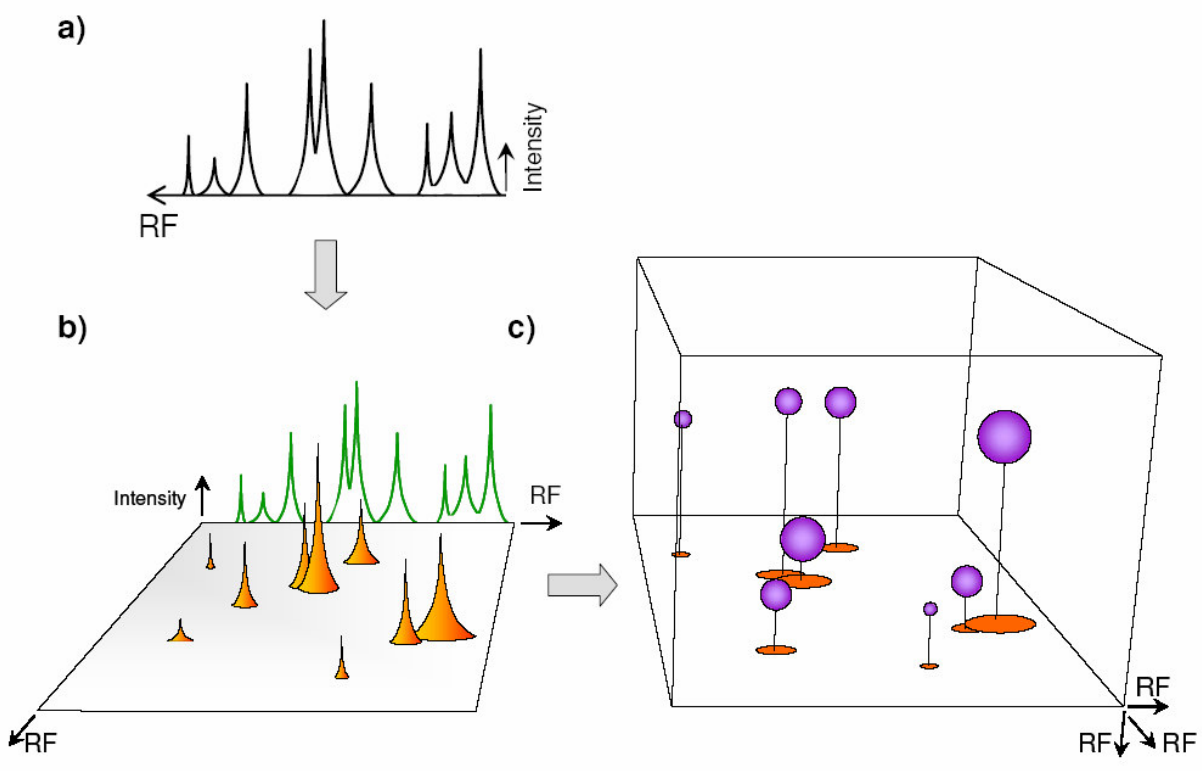


Fig. 5: Different techniques can be used to spread a one dimensional spectrum (a) over 2 dimensions (b), three dimensions (c) or more. In a), the resonance frequency (RF) is on the horizontal axis and the intensity on the vertical axis. In b) the resonance frequencies are on the horizontal axes and the intensity on the vertical axis. Peaks appear as cones (in orange, the original 1D spectra is shown in green). In c), the resonance frequencies are along the three axes, and peaks appear as spheres which diameter is proportional to the intensity (in violet, the original 2D peaks being shown in orange). Note how peak overlap is reduced by increasing the number of dimensions.

2D and 3D NMR techniques have been developed. In such spectra, e a 1D spectrum is spread out over one or more extra dimension: instead of having n peaks along one axis, those n peaks are scattered over a surface or in a volume (fig. 5).

4. 2D NMR can be used to observe structural changes in a protein

A very useful type of 2D NMR experiment in protein sciences is the HSQC (Heteronuclear Single Quantum Coherence) experiment. This experiment requires all nitrogen atoms in the sample protein to be ^{15}N.

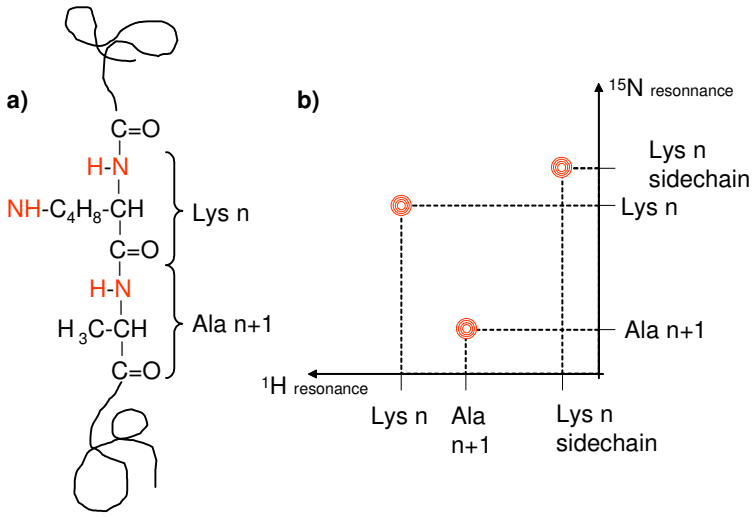


Fig. 6: a) Two amino acids of a protein with their amide groups highlighted in red. b) the peaks obtained for those amide groups in a HSQC spectrum in the usual "level curves" representation. Real HSQC spectra are presented in Fig.7

The way the sample is irradiated (pulse sequence) allows to observe the resonance frequencies of hydrogen and nitrogen atoms involved in N-H bonds. In the resulting HSQC spectrum, a NH group will be observed as a peak at the resonance frequencies of its Hydrogen and Nitrogen atoms (fig. 6).

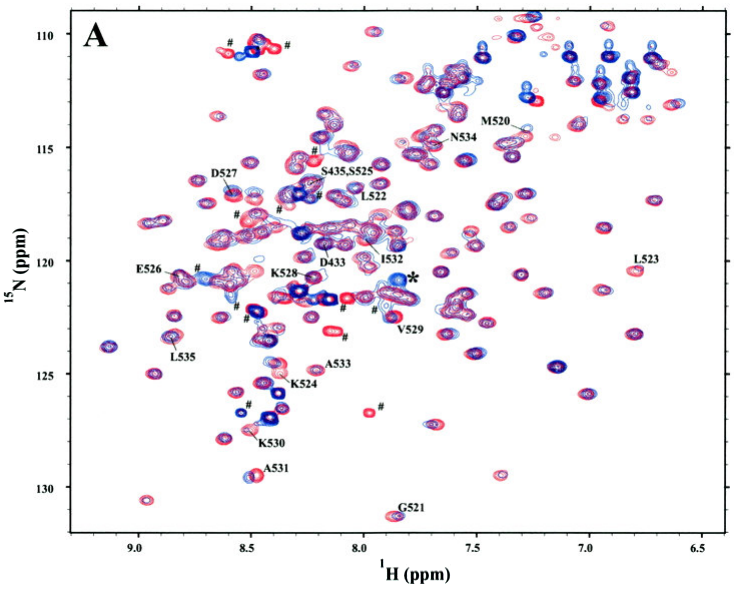


Fig. 7: Example of use of HSQC Spectra for investigating the effect of conditions changes on a protein. In this case, spectra of non sumoylated (red) and Sumoylated (blue) RanGap1 have been overlaid. The limited extent of the differences between both shows that Sumoylation has little effect on the overall structure of RanGap1 (see also the section "interactions of SUMO with other proteins).

Such a spectrum of a protein can be measured in few hours (the higher the concentration, the less measuring time is needed to generate data of adequate quality). Alone, a HSQC spectrum of a protein is of little interest. It gives an idea of the α-Helix and β-strand content of a protein, or of whether the protein is aggregated, but little else. The great

interest of HSQC spectra is for monitoring structural changes in a protein. Changes in the structure of a protein mean that the relative position of its atoms and their electronic clouds change as well. This results in changes in the local magnetic fields experienced by the various nuclei. Hence, structural changes can be seen in NMR as shifts of the resonance frequencies of the nuclei of a protein. In HSQC, these are observed as changes of the position of the corresponding peaks. Binding of a ligand to a protein can be monitored by this method because it changes the local magnetic field of the protein directly –as any molecule, a ligand generates a magnetic field- and indirectly by inducing structural changes in the protein (fig. 7).

5. 3D NMR and the assignment of the resonances of a protein

We have seen in the previous section that 2D NMR spectra can be used to investigate structural changes in proteins, by observing the effect of condition changes on the resonance frequencies of Hydrogen and Nitrogen nuclei. More detailed information is available in such spectrum if one is able to determine to which amino acid of the protein corresponds each peak in the spectrum. Unfortunately, such information cannot be derived from the HSQC spectra themselves, and other

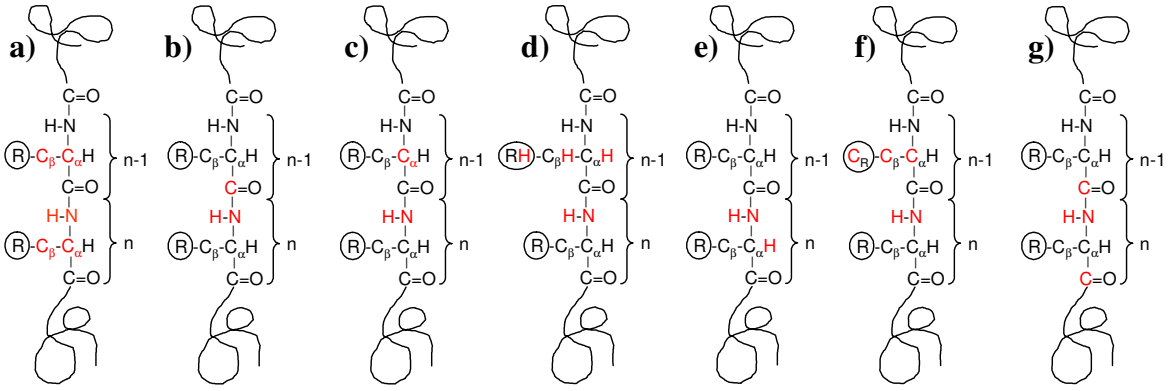


Fig. 8: atoms visible in different types of spectra for a generic dipeptide within a 15N 13C marked protein. Note that the projection of each of those spectra (except the HN(CO)CA) onto the 1H15N plan amounts to a HSQC spectrum. a) HNCACB b) HNCO c) HN(CO)CA d) HC(CO)NH e) HNHA f) C(CO)NH g) (HCA)CO(CA)NH.

types of spectra have to be analyzed to assign the observed resonance to the amino acids they come from. In the “good old days” – a mere decade ago- two dimensional spectra were used. Technical progress has brought about new methods and better

performing spectrometers to apply them, and the easiest way to assign resonances to their amino acids is to use a set of three dimensional spectra.

The productions of those spectra require that all the carbons in the protein are ^{13}C and all the nitrogens ^{15}N. The resulting spectra have 3 axes, one for the Carbon resonances, one for the nitrogen resonances and one for the hydrogen resonances. By using different pulse sequences, one can select the atoms that will be observed, as shown in fig. 8.

Ideally the assignment of the protein backbone can be produced using only a HNCACB spectrum. In this spectrum, peaks come in quartets, as shown on fig. 9. A quartet is composed of four peaks that have the same N and H resonances. Those resonances correspond to the amine group of the n^{th} amino acid in the sequence. The two big peaks of the quartet correspond to the resonances of the C_α and C_β atoms of the amino acid, the two smaller peaks to the resonances of the C_α and C_β atoms of the (n-1^{th}) amino acid. As the usual resonance ranges for the C_α and C_β resonance frequencies of the various amino acids are known, it is possible to make hypothesis on the nature of each the two amino acids observed in a quartet of peaks. Generally, there are less than half a dozen of possibilities for a pair of peaks. This gives a certain number of possible combinations for a quartet of peaks. However, most of those combinations are not found in the sequence of the studied protein, and usually one is left with one or two assignment alternatives for a quartet. Quartets can also be sequentially ordered, taking advantage that the amino acid n-1 is observed giving peaks at the same ^{13}C coordinates in two quartets: it corresponds to the two small peaks in the [n; n-1]

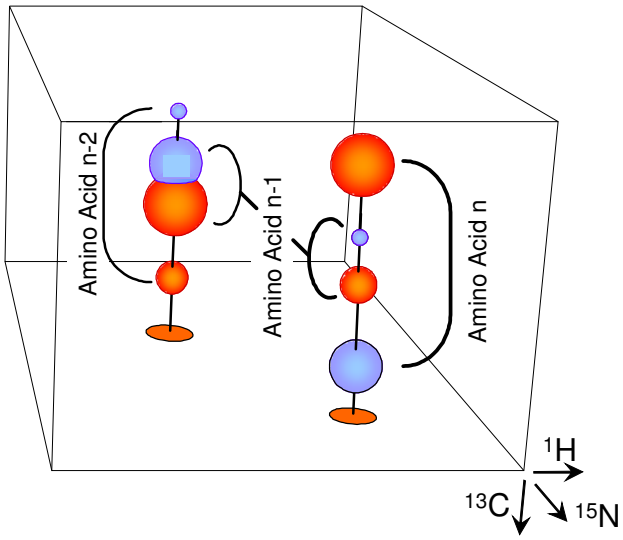


Fig. 9: two "quartets" of peaks in an HNCACB spectrum. The Cα and Cβ of an amino acid give peaks of similar intensity with opposite signs (color coded as red and blue). The amino acid number n-1 appears twice in the spectrum, once as the two bigger peaks of a quartet, and once as the two smaller ones in another quartet. Its Cα and Cβ having the same coordinates in both quartets, the quartets can be sequencially ordered.

quartet, and to the two big peaks in the [n-1; n-2] quartet (fig. 9). Other types of 3D spectra may be used when there are ambiguities or lacunes in the HNCACB spectrum, and, when the assignment of the protein backbone is known, to assign the resonances of the atoms that are not observed in the HNCACB spectrum (i.e. the carbon in the C=O group of the peptide bond and the Carbon, Nitrogen and Hydrogen atoms in the side chains of the amino acids).

II) SUMOylation

1. Post-translational Modifications

When the human genome was published in February 2001 [International Genome Sequencing Consortium, 2001], the greatest surprise came from the number of genes found in it. The news that a few tens of thousands of proteins –ten time less than what was usually expected- suffice to assure the functions required for perpetuating human life was welcomed by general incredulity: how could the combination of so few individual functions result in such a complex organism as a human being? A Jumbo Jet is assembled from some 200000 types of parts, and is comparatively extremely simple.

One has, however, to consider that, for various reasons a gene usually codes for more than one function. Each gene may code for more than one protein (due to alternative splicing). Most proteins are composed of several domains, each having one or more functions. Furthermore, each protein can be, in the course of its life, subjected to post-translational modifications in response to intracellular events or changes in extracellular conditions.

Post-translational modification is defined a physiological change in the arrangement of covalent bonds in a protein, without counting the chemical bonds formations implied by enzymatic mechanisms. This includes cross linking reactions between the amino acids of a protein (as in GFP and proteins containing disulfide bridges), the removal of parts of a protein (as in insulin) and the covalent binding of a chemical compound to a protein. Protein constitutively containing a non covalently bound chemical compound required for their function and structural integrity (as hemoglobin with haem or rhodobsin with carotene) may be regarded as post-translationally modified as well. The possible chemical groups are various in nature. Some, as phosphate, are very small and inorganic whereas others (SUMO, Glycosylation…) are large organic molecules. Most post-translational modifications are carried out by enzymes which recognize a specific motif in a protein. Some post-translational modifications (a good example is the Glycosylation of cartilage proteins) are constitutive, meaning that all the concerned proteins bear those modifications whatever the context. However the majority of post-translational modifications are carried out in response to changes in the intra- or extracellular environment.

The role of post-translational modifications is to change the affinity of a protein to chemical compounds. This results in changes in localization (e.g. some proteins are attached to a lipid and thus become anchored to a membrane), enzymatic efficiency, ability to recognize interaction partners…

The recognition of the importance of phosphorylation has brought the studies of post-translational modifications to emerge as a major field of biological research since over a decade.

2. Description of SUMO proteins

SUMO (Small ubiquitin like modifier) proteins have been discovered in the nineties by [Meluh & Koshland, 1995]. They have received their name because like Ubiquitin – and several protein discovered since then- they can be covalently attached to target proteins through an isopeptide bond (fig. 10).

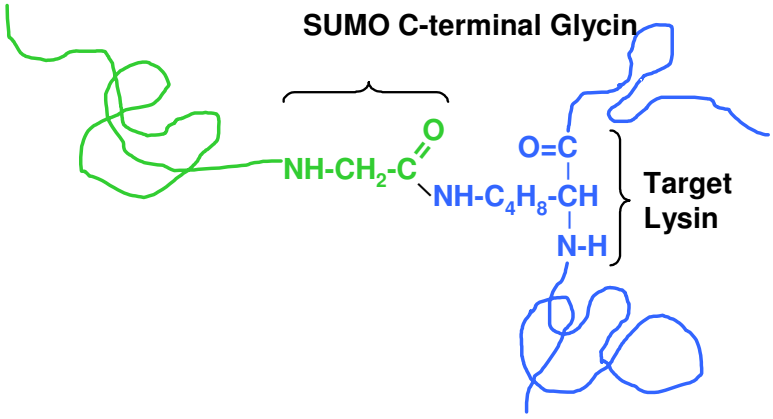


Fig. 10: Sumoylated protein. The SUMO moiety is in green, the target protein in blue.

SUMO is however not the only name given to those proteins, and the

Name in the present work	N-terminal amino acids	alternative name	database	ID	Original publication	Name in original publication	Gene sequence	database	PDB ID of structure
SUMO1	MSDQEA...	SMT3-C_HUMAN	SwissProt	P63165	Lapenta V. *et al*, 1997	SMT3C	X99586	Entrez Nucleotides (NCBI)	1A5R
		SUMO1	RefSeq (NCBI)	NP_003343			NM_003352.4	RefSeq (NCBI)	
		PIC1, Ubl1, SMT3H3, GMP1, Sentrin							
SUMO2	MSEE...	SMT3-A_HUMAN	SwissProt	P55854	Lapenta V. *et al*, 1997	SMT3A	X99584	Entrez Nucleotides (NCBI)	1U4A, 1WM3
		SUMO3	RefSeq (NCBI)	NP_008867			NM_006936	RefSeq (NCBI)	
		SMT3H1							
SUMO3	MADEK...	SMT3-B_HUMAN	SwissProt		Mannen *et al*, 1996	MIF2 Suppressor	L76416	Entrez Nucleotides (NCBI)	1U4A, 1WM3
		SUMO2	RefSeq (NCBI)	NP_008868			NM_006937	RefSeq (NCBI)	
		SMT3H2, Sentrin2							

Table 1: SUMO isoforms in the databases

numbering of the isoforms differs from one author to the other, which calls for caution when comparing data from different sources. The different nomenclatures and databases entries are summarized in table 1.

In man, three SUMO isoforms have been identified [for SUMO1 and -3, Lapenta *et al*, 1997; for SUMO2, Mannen *et al*, 1996]. A fourth has been reported in summer 2004 [Bohren *et al*, 2004], but whether this isoform actually exists is still the subject of heated discussions.

SUMO2 and -3 are more closely related to each other (96% identity) than to SUMO1 (~45% identity in both cases) (fig. 11). The structure of SUMO1 and -3 has been solved by different groups [Bayer 1998; Huang *et al*, 2004]. As the structured region of SUMO3 is identical to the structured region of SUMO2, it can be considered

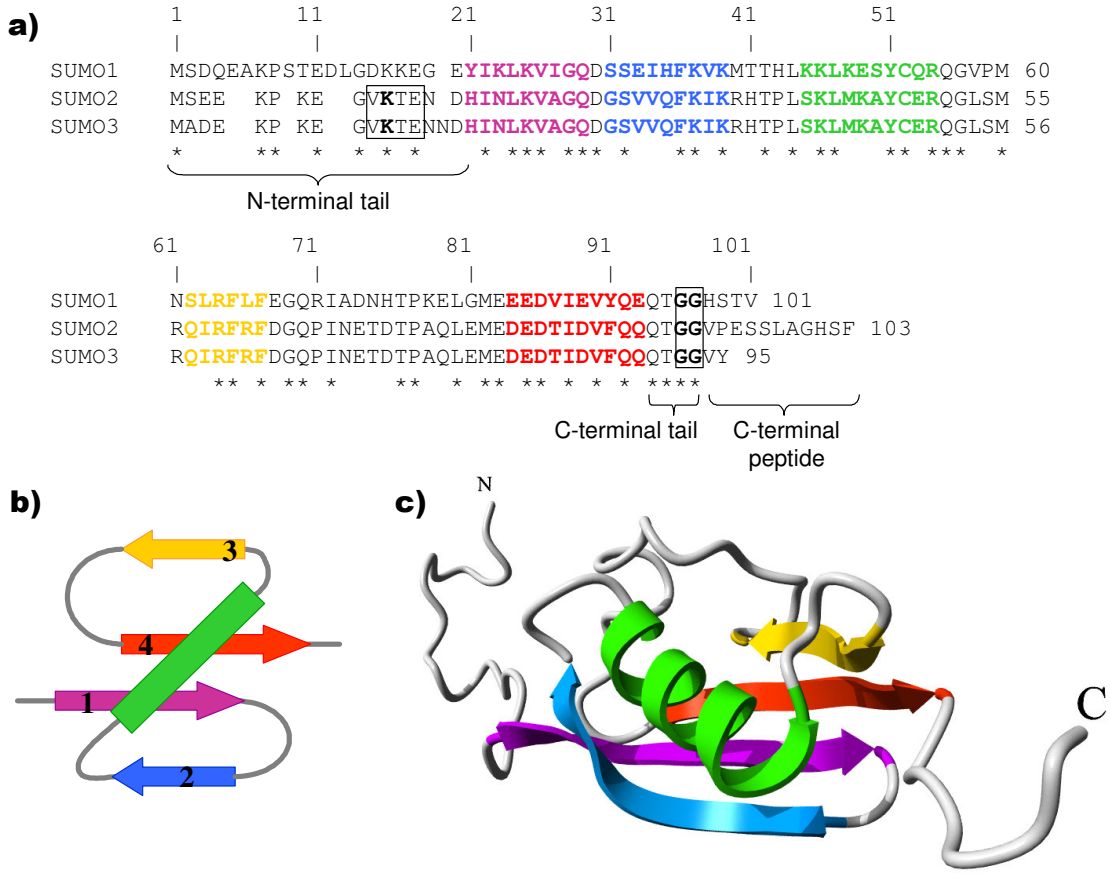


Fig. 11: Features of SUMO proteins. a) Alignment of human SUMO 1, 2 and 3. Structural elements are highlighted in the same colors as in b) and c). The Sumoylation motif in the N-terminal tail of SUMO2 and 3 is boxed, with the target lysin highlighted in bold. The Gly-Gly motif in the C-terminal tail is boxed as well. b) Diagram of the fold common to SUMO and Ubiquitin. c) Structure of SUMO1 (PDB ID: 1A5R, Bayer *et al*, 1998)

that the structure of SUMO2 is known as well.

Despite the low sequence homology between them, SUMO and Ubiquitin have very similar tertiary structures. This structure consists in a (-sheet made of 4

strands and a (-helix (fig 11b and 11c). Unlike Ubiquitin, SUMO proteins have long unfolded C- and N- termini ("tails") in which most of the differences between the isoforms are concentrated.

All SUMO have in their C-terminal sequence two consecutive glycine residues ("Gly-Gly motif") followed by a "C-terminal peptide" which is cut off at the beginning of the Sumoylation process (fig 11a).

SUMO2 and -3, but not SUMO1 have a Sumoylation site (i.e. they can be sumoylated at that place) in their N-terminal tail (fig 11a), allowing the formation of poly-SUMO chains.

SUMOylation has very various and numerous functions in many cellular processes (among other, transcription control, regulation of ubiquitinylation, transcription controls…). Those functions have been often reviewed over the last decade, and particularly well so by [Dohmen, 2004]. This section will therefore not detail those aspects and concentrate on the structural knowledge on SUMO, which has not yet been the subject of a dedicated review.

3. The Sumoylation process

Sumoylation is carried out by a machinery similar to that responsible for ubiquitinylation. It is a complex process involving four steps. Due to the effort of several research groups, we now have a good structural knowledge of those steps, which will be presented in the present section.

The first step –removal of the amino acids following the Gly-Gly motif- is performed by several dedicated proteases. Each of those proteases is able to process any SUMO isoforms, albeit with very different affinities. In practice, each SUMO isoforms appears to be processed by a particular SUMO protease. Substrate recognition by SUMO proteases is exclusively based on the sequence of the C-terminal peptide.

Since the enzymes catalyzing the following step of Sumoylation specifically recognize a C-terminal Gly-Gly motif on SUMO as a substrate, it is essential that the removal of the C-terminal peptide takes place. Failure of the SUMO proteases to perform their function leads to absolute inhibition of Sumoylation.

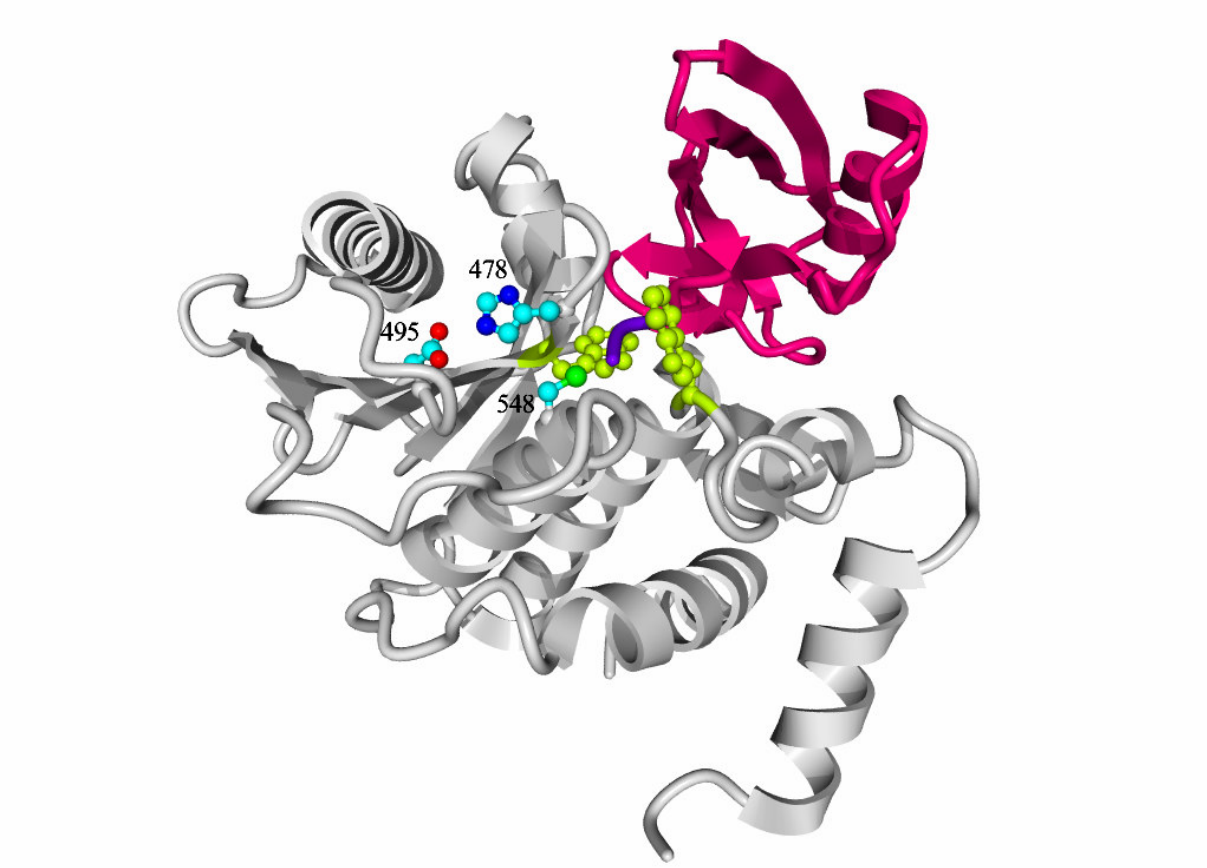


Fig. 12: Structure of the SUMO1-Senp2 complex. SUMO1 is in pink, Senp2 in light grey. The side chains of the active site residues are in CPK colors, and the two tryptophan residues forming the tunnel through which the Gly-Gly motif accesses the active site are in lime green.

Structures of the SUMO protease Senp2 alone (1THO) and in complex with SUMO1 (1TGZ) have been solved [Reverter & Lima, 2004] have been solved. The latter structure is of a complex artificially blocked after the protease reaction has taken place, but with the Gly-Gly motif still in the active site. This structure shows why it is necessary for the function of the enzyme that Gly-Gly motif is present in C-terminal of SUMO: the “tunnel” leading to the active site is too narrow for any other amino acid to fit into it (Fig. 12). Since the C-terminal peptide is not present in this structure, we still have no structural information on the mechanism by which Senp2 recognizes SUMO isoforms. The contact surface of SUMO1 for Senp2 is shown in Fig. 13.

In the second step of Sumoylation, SUMO is transferred to an E1 enzyme (in man, the Sae1/Sae2 complex). Sae2 adenylates the C-terminal Glycin of SUMO and then forms a thioester bond between this Glycine and a Cysteine of Sae2. The energy gained in the hydrolysis of the ATP is stored in the thioester bond to be finally used in the Sumoylation reaction. Structures of the Sae1/Sae2 complex alone and with SUMO1 (PDB entries 1Y8Q and 1Y8R respectively) have been published recently

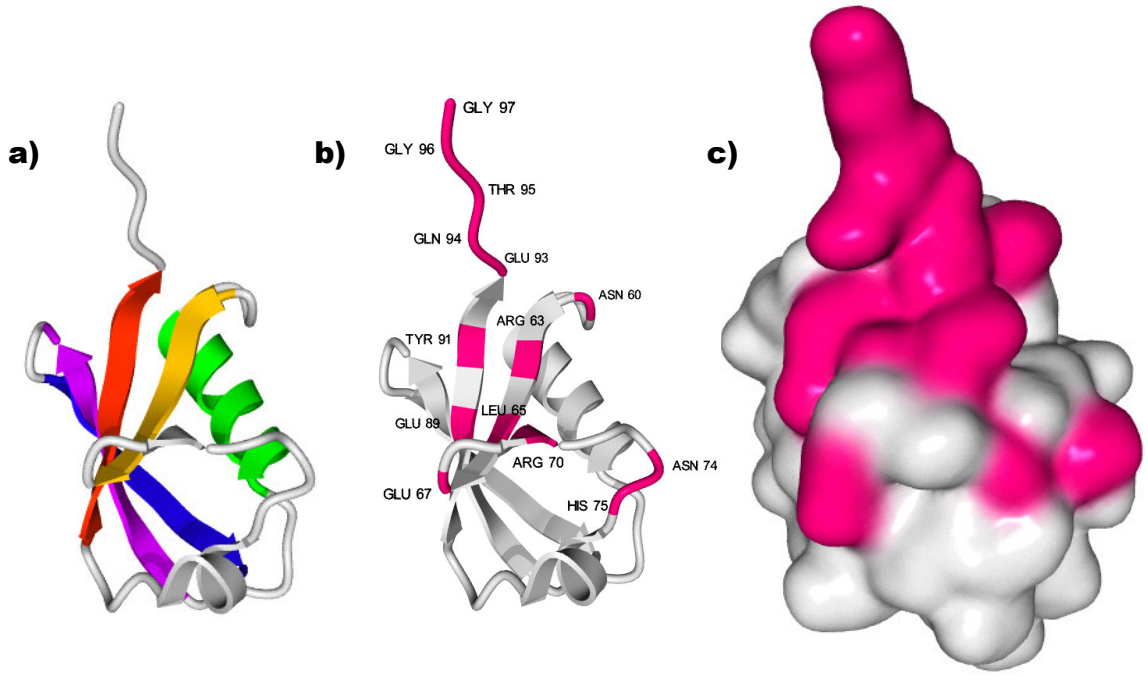


Fig. 13: Interaction surfacor Senp2 on SUMO1. a) Structure of SUMO1 with the same color convention as on fig. 2. b) Structure of SUMO1 with the same orientation as in a), amino acids contacting Senp2 are in pink. C) Surface representation of SUMO1, amino acids contacting Senp2 are in pink, the orientation is the same than in a) and b). Amino acids of a protein are considered to contact another protein when at least one of their atoms is less than 7Å away from an atom of that protein. The program Protein Explorer was used to find such contacts in all the protein complexes presented in this and the following figures.

[Lois & Lima, 2005] (fig. 14). It shows that though yeast Sae1 is essential for Sumoylation to take place, it has neither direct implication in the E1 reaction, nor direct interaction with SUMO. Its role is to hold Sae2 in its correct conformation. The Sae2 moiety performs all the tasks required for the E1 function. It is made of three domains: the Cysteine Domain, the Adenylation domain and the Ubl domain. They delimitate a wide cavity that receives the C-terminal face of SUMO. SUMO's C-terminal tail extends in a crevice on the adenylation domain surface leading to the active site where the C-terminal glycine of SUMO contacts ATP. In the following intermediates, which are not represented in the published structure, the C-terminal Glycine is adenylated, and conformational rearrangements in the E1-SUMO complex bring it in proximity of an active Cysteine in the Cysteine domain, to which it becomes bound through a thioester bond. The Ubl domain recruits the E2 enzyme Ubc9 for the following Sumoylation step, but no structural data is available on this interaction. The contact surface of SUMO1 for Sae1/Sae2 is shown in Fig. 15.

In the third step of Sumoylation, SUMO is released from Sae2 and its C-terminal Glycine forms a Thioester bond with a Cysteine of the E2 enzyme Ubc9. Ubc9 also recruits the target protein. Several structural studies have investigated this

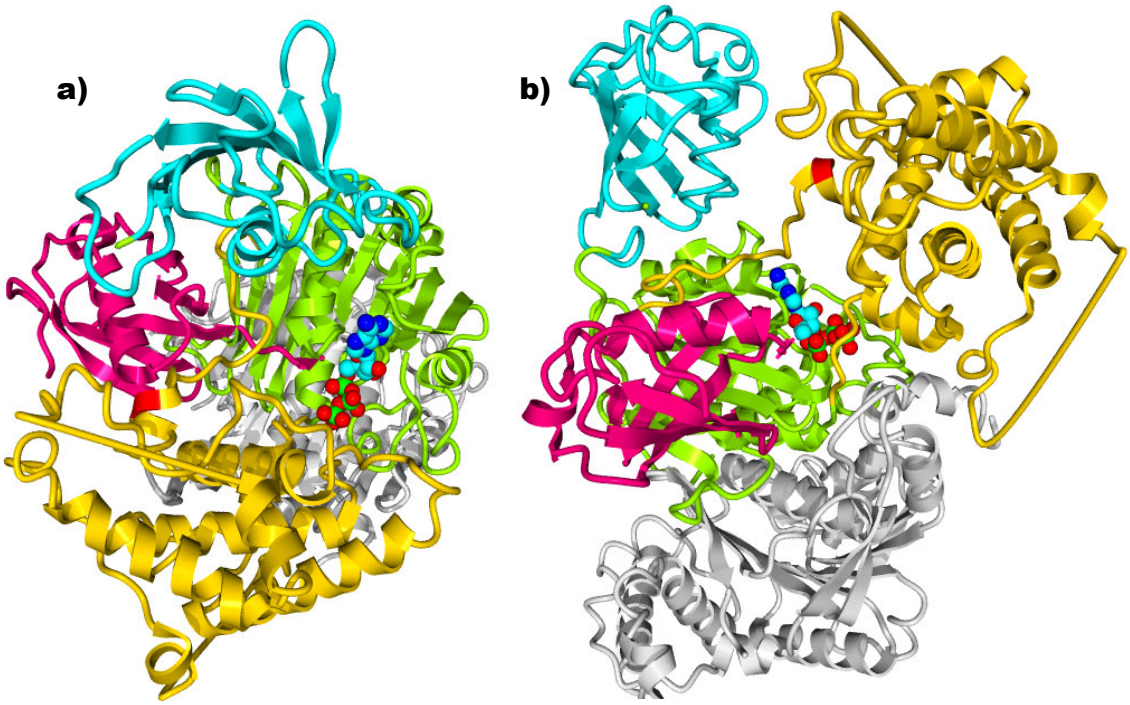


Fig. 14: Sae1/Sae2-SUMO1 complex viewed from two different points (a and b). SUMO1 is in pink. Sae1 is in grey. Sae2 domains are colored in different colors: the Ubiquitin-like domain is in blue, the adenylation domain in green and the Cystein domain in yellow. The position of the cystein to which the C-terminal Glycin of SUMO becomes attached is colored in red (in the present structure, it has been mutated to an alanine). The ATP molecule is in balls and sticks and CPK colors.

step. The Structure of Ubc9 (1U9A) has first been solved by [Tong *et al*, 1997]. The interactions between SUMO and Ubc9 have been studied by NMR [Liu *et al*, 1999;

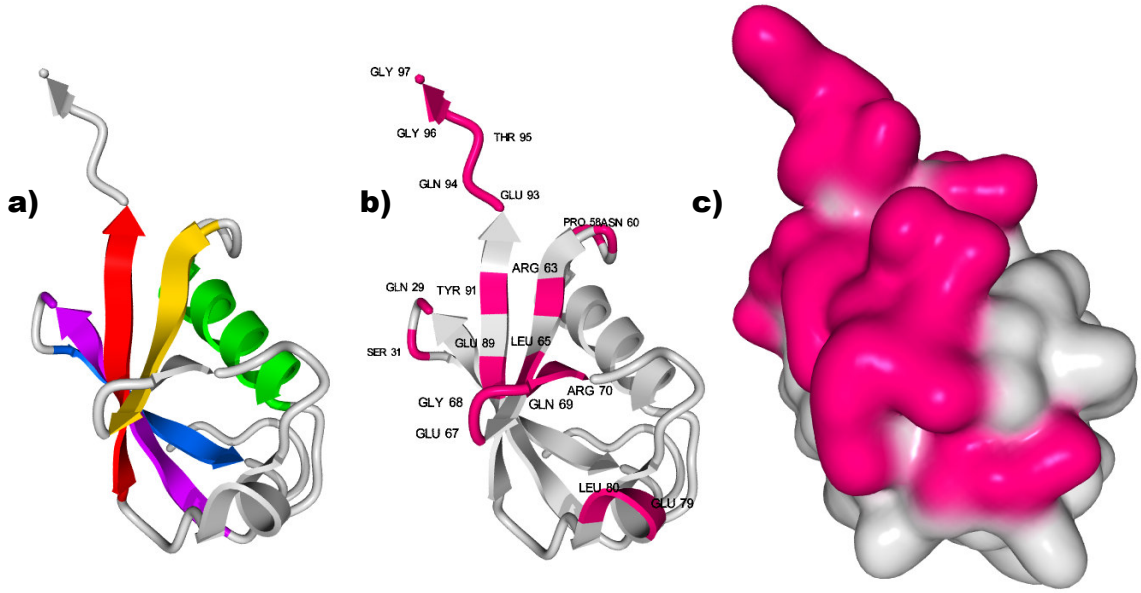


Fig. 15: Interaction surface for Sae1/Sae2 on SUMO1. a) Structure of SUMO1 with the same color convention as in fig. 2. b) Structure of SUMO1 with the same orientation as in a), amino acids contacting Sae2 are in pink. c) Surface representation of SUMO1, amino acids contacting Sae2 are in pink, orientation is the same as in the previous panels.

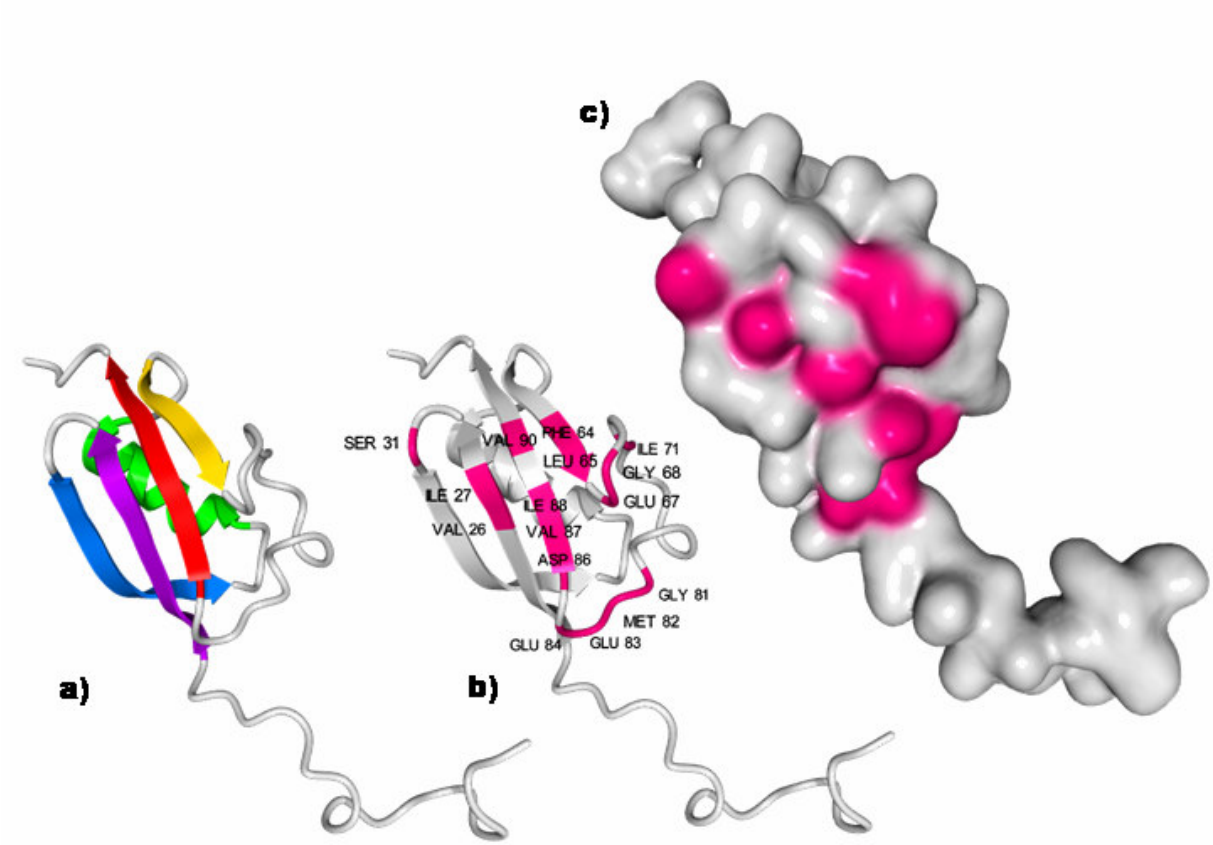


Fig. 16: Interaction surface for Ubc9 on SUMO1. a) Structure of SUMO1 with the same color convention as in fig. 2. b) Structure of SUMO1 with the same orientation as in a), amino acids contacting Ubc9 are in pink. c) Surface representation of SUMO1, amino acids contacting Ubc9 are in pink, orientation is the same as in the previous panels.

Tatham *et al*, 2003]. Those works show that SUMO contacts Ubc9 through its β-sheet, which offers a wide negatively charged interface [Bayer *et al*, 1998] (fig. 16). This surface contacts the area of Ubc9 surface defined by its N-terminal α-helix and first β-strand (fig. 17). This contact is important for the formation of the thioester bound SUMO1-Ubc9 complex, but once this complex is formed, it appears to be non essential for the Sumoylation reaction [Tatham *et al*, 2003]. The opposite face of Ubc9 recruits the target protein (fig. 17). The Ubc9-target interface on Ubc9 has been observed for Ubc9 interactions with p53 and c-Jun [Lin *et al*, 2002] and RanGap1 [Bernier-Villamor *et al*, 2002]. The structural determinants that cause the target to be recognized by Ubc9 are not identified. The active site of Ubc9 is placed between the binding sites for SUMO and the target protein. The target lysine and the C-terminal tail of SUMO extend over the surface of Ubc9 to reach it (fig. 17).

The fourth step of Sumoylation is the Sumoylation reaction itself. It is calalysed by Ubc9 and an E3 enzyme. In contrast to the other Sumoylation enzymes seen previously, E3 enzymes are neither unique (at least four are known: PIAS3, PIASxα,

RanBP2 and TOPORS) nor essential for the Sumoylation to take place. The role of the E3 enzyme is generally regarded as merely structural: it should maintain SUMO and Ubc9 in a favorable relative orientation but does not make any covalent bond with SUMO or the target protein (the Sae1 protein discussed earlier in this section has a similar role). Recent studies have shed some light on the structural aspects of the interactions between SUMO1 and PIAS [Song *et al*, 2004] and SUMO1 and RanBP2 [Reverter & Lima, 2005]. The later shows that RanBP2 a few amino acids of RanBP2 form an intermolecular β-sheet along the 2nd β-strand of SUMO1. The former shows that PIAS interacts with a region centered on the same β-strand of

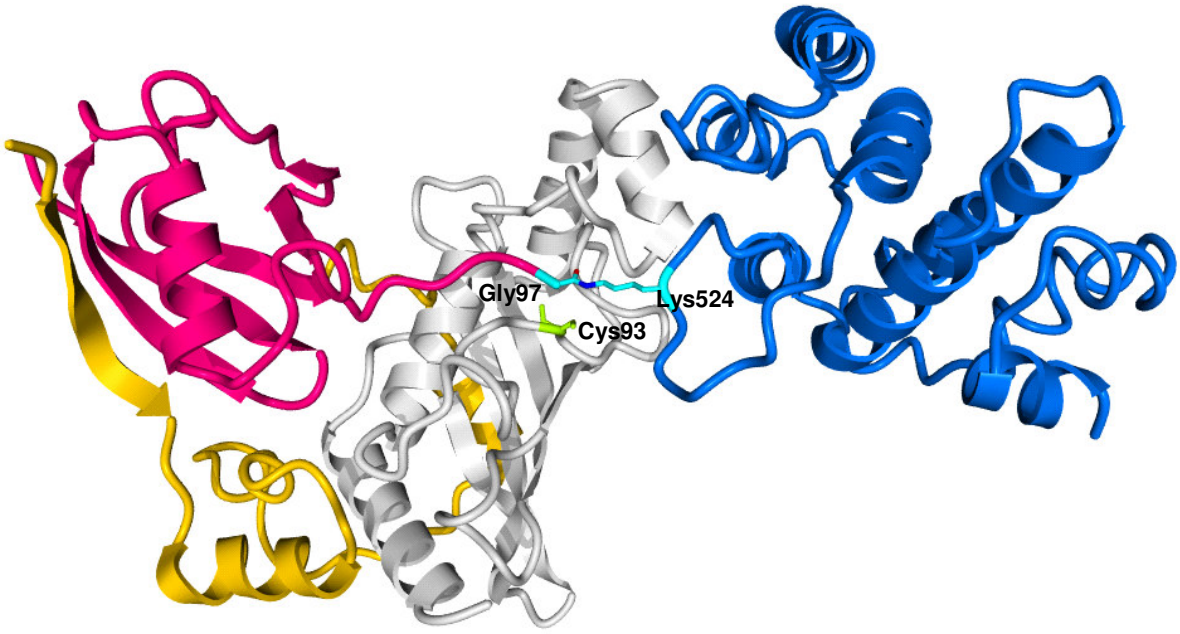


Fig. 17: The final Sumoylation complex. SUMO1 is in pink, RanGap1, the target protein, in blue, Ubc9 in grey and RanBP2, the E3 protein, in yellow. The side chain of the catalytic cystein (residue 93) of Ubc9 is in lime green, the side chains of Glycine 97 of SUMO1 and of the target lysine 524 of RanGap1 are represented in CPK colors.

SUMO1. The interaction surface of Ubc9 for RanBP2 has been found by [Tatham *et al*, 2005] to be the face of Ubc9 β-sheet opposite to its active site.

4. Interactions of SUMO with other proteins

As we have seen in the previous section, most of the SUMO interactions involved in Sumoylation are well known. The same is not true for the interactions between SUMO and proteins that do not take part in The Sumoylation process.

Those proteins can be divided into two classes: target proteins and non-target proteins. A better knowledge of those interactions is essential to understand how SUMO proteins do function.

There has been long discussion on whether SUMO functions as a mere tag that recruit some interaction partners that the unsumoylated target protein couldn't recruit; by inducing structural changes in the target protein; or by preventing access

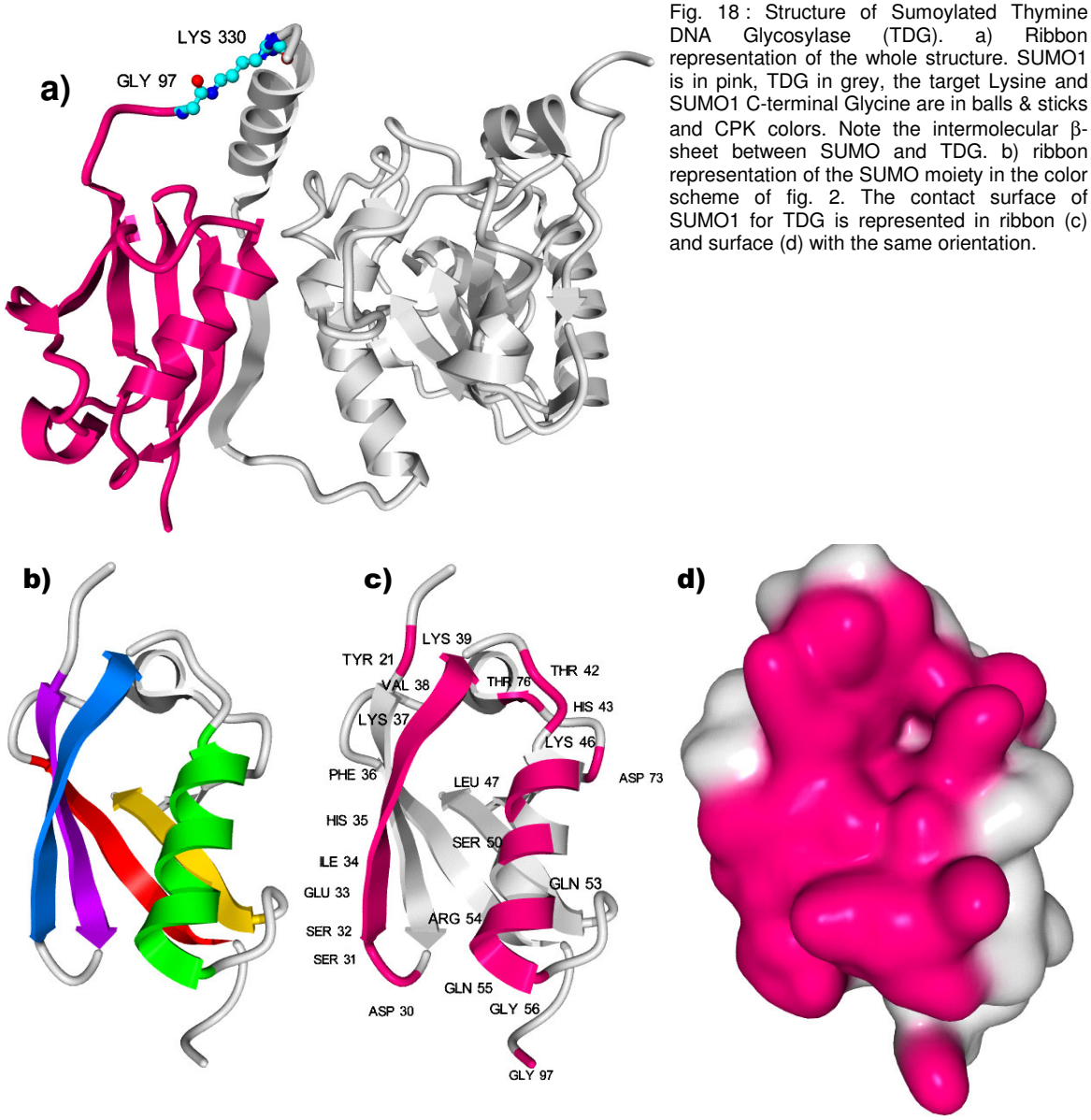


Fig. 18 : Structure of Sumoylated Thymine DNA Glycosylase (TDG). a) Ribbon representation of the whole structure. SUMO1 is in pink, TDG in grey, the target Lysine and SUMO1 C-terminal Glycine are in balls & sticks and CPK colors. Note the intermolecular β-sheet between SUMO and TDG. b) ribbon representation of the SUMO moiety in the color scheme of fig. 2. The contact surface of SUMO1 for TDG is represented in ribbon (c) and surface (d) with the same orientation.

to some part of the target protein's surface. The past monthes have seen the publication of three structures examplifying each of those three possibilities. The two latter used to be the most popular, until a recent study on Sumoylated RanGap1 [mac Auley *et al*, 2004] showed that there are no interactions between the SUMO and

RanGap1 moieties in Sumoylated RanGap1: they simply behave like a tether. However, another study showed that target proteins do not always have such freedom toward their SUMO tag: the structure of Sumoylated Thymine DNA Glycosylase (TDG) (PDB entry 1WYW) [Baba *et al*, 2005] shows that both moieties are entangled through an intermolecular β-sheet (fig. 18). This structure allows to understand how Sumoylation is crucial for TDG to accomplish its function. No-Sumoylated TDG binds to DNA and recognizes mismatched DNA nucleotides and excises a base from the mismatched pair. This results in conformational changes in TDG that cause it to clash with and release the bound DNA, letting the repair site free for an AP endonuclease which performs the following step in the repair process.

The structure of the Ubiquitin E2 enzyme E2-25K published by [Pichler *et al*, 2005] (fig. 19) show that SUMOylation of the N-terminus of this protein does not substancially modify its structure. Instead, the SUMOylation site of E2-25K is placed in its E1 binding site. Thus, the SUMO moiety of SUMOylated E2-25K plays its

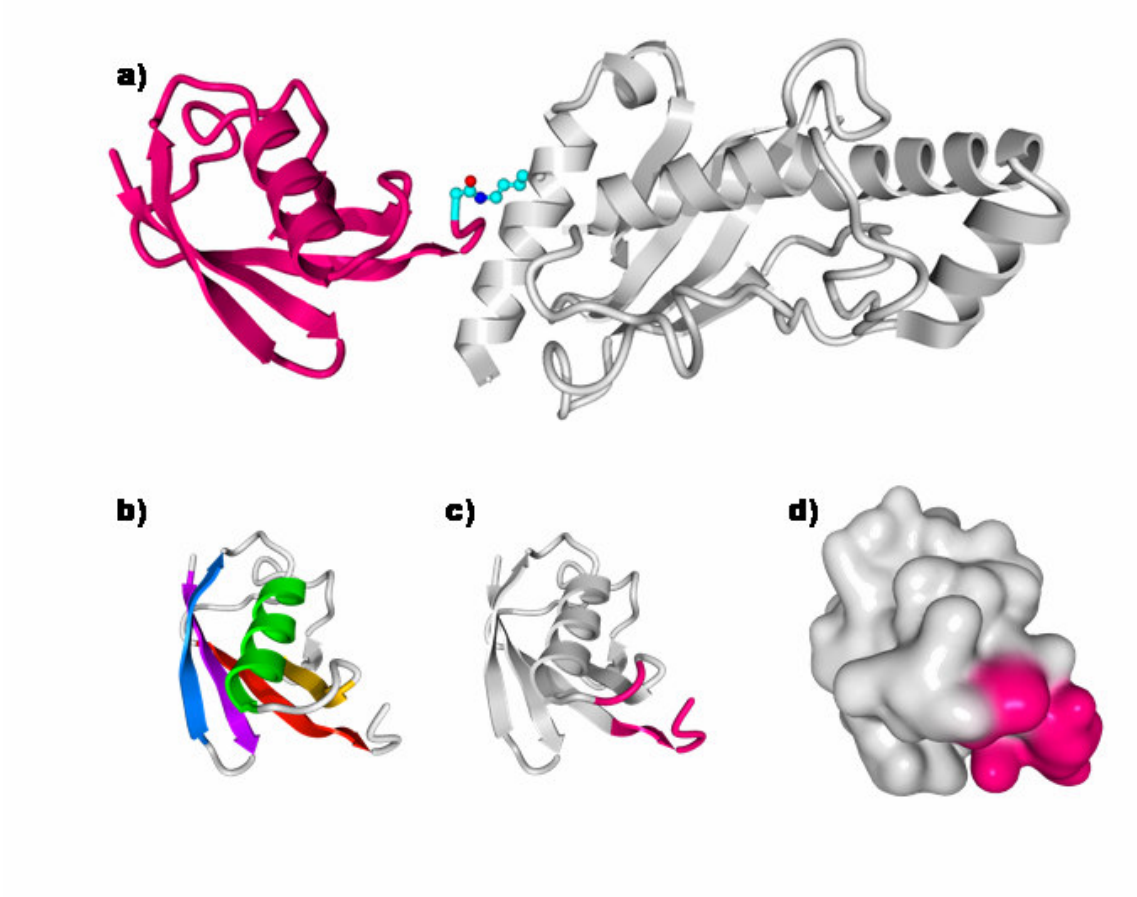


Fig. 19: Structure of Sumoylated E2-25K. a) Ribbon representation of the whole structure. SUMO1 is in pink, E2-25K in light grey, the target Lysine and SUMO1 C-terminal Glycine are in balls & sticks and CPK colors. b) ribbon representation of the SUMO moiety in the color scheme of fig. 2. The contact surface of SUMO1 E2-25K is represented in ribbon (c) and surface (d) with the same orientation.

inhibitory role by hindering the access of the E1 enzyme to its binding site on E2-25K.

If little is known on the structural basis of the interactions between SUMO and their target proteins, even less has been described on the interactions between SUMO other proteins. An interaction surface on SUMO1 for a $PIAS_{x\alpha}$ derived peptide has been determined by [Song *et al*, 2004]: it consists in amino acids of the second β-strand and the consecutive α-helix. However, the question of how the selectivity between the SUMO isoforms is achieved remains open: have some proteins more affinity for a particular SUMO isoform, and if so, what are the structural determinants for this selectivity? The present work is part of a cooperation project aiming at answering those questions by biochemical and biophysical means.

III) Identification of a SUMO-interacting motif in non-target proteins

1. Molecular biology of SUMO interactions

Several studies have been dedicated to the definition of a "SUMO interactome" [Hannisch *et al*, 2005, Song *et al*, 2004], and have concentrated on identifying proteins interacting with SUMO1. The present work is part of a project aimed at comparing the interactions with the different SUMO isoforms. Since the results of the biochemical experiments made in this project are the basis of the work presented in this thesis, they will be presented shortly in this section.

Yeast Two Hybrids experiments were used to determine proteins binding to SUMO1 or 2. In those experiments, SUMO isoforms lacking the C-terminal Gly-Gly motif were fused to the Gal4 DNA binding domain of YTH9 and used as bait. The deletion of the Gly-Gly motif prevents the bait SUMO from being covalently attached to target proteins, ensuring that only proteins directly interacting with SUMO can be identified by those experiments. As a prey, cDNA banks from Spleen, Thymus and Kidney were used. The strength of the interaction between SUMO and the various preys was estimated by the intensity of the coloration obtained in X-Gal test. 21 SUMO interacting proteins could be identified, 10 of which had not been identified as SUMO interacting proteins previously. With exception of previously identified Sumoylation proteins, the identified proteins all have functions involving DNA or RNA: transcription factors, helicases, repair enzymes, etc… Most of those interact equally well with SUMO1 and 2. Some, however, showed preference for either isoform. Those results were confirmed by GST Pulldown experiments.

In most cases, the SUMO interacting motifs identified in previous studies could be identified in the proteins retrieved from the Yeast Two Hybrid experiments. However, this was not the case for three of them (namely TTRAP, MCAF and ZCCHC12). TTRAP was selected as representative from those three proteins. It could be shown by GST pulldown experiments that exclusively the C-terminal part of TTRAP has ability to bind SUMO. It contains the sequence at Ile-Val-Asp-Val positions 280 to 284, which corresponds to the SUMO interacting motif defined by [Song *et al*, 2004] read backwards. To confirm that those amino acids are indeed responsible for binding to SUMO, half or all of them were mutated to alanine. The resulting mutant had very limited or inexistent capacity of binding to SUMO,

confirming the essential role of the Ile-Val-Asp-Val tetrapeptide of TTRAP in SUMO binding.

Bioinformatic analysis of the sequence of the identified SUMO interacting proteins enabled to define a SUMO Interacting Motif (SIM). It consists of a hydrophobic core made of 3 Isoleucines or valines and another amino acid preceded or followed by a negatively charged tract of around 15 amino acids. In most cases, several serines are found between the hydrophobic core and the negative tract. Those serines, as will be discussed later, are a phosphorylation site. Thus, this new definition of the SIM is also a synthesis of the elements formerly shown to be important for SUMO binding.

2. Aim of the present studies

Having defined a short motif responsible for SUMO binding in a large number of proteins clamored for understanding the modalities of this binding and the mechanism by which SUMO isoform specificity can be achieved by the different SIM despite their strong overall similarity. The aim of the present work is to provide structural understanding for the binding of the SIM to SUMO1 and SUMO2 and the ways SUMO isoforms specificity is achieved: what kind of contacts between the SIM and SUMO are involved, and what are the roles of the different elements of the SIM, and how the SIM discriminates between different SUMO isoforms. Since it allows observation of the behavior of single atoms in a molecule in solution, NMR was chosen as the method to observe the effect of the binding of peptides containing all or parts of the SIM of PIAS and TTRAP on each individual amino acids of SUMO.

Materials & Methods

1. Cloning, Protein expression and purification for NMR studies

Full length SUMO1 and SUMO2 –preceded by a thrombin cutting site- were cloned as GST-fusions into pET-41a vectors (Novagen). The original thrombin cutting site of this vector was mutated using the QuickChange mutagenesis kit (Stratagene) from according to manufacturer's instructions with the following primers *pET41 delete THRB fwd* (cg ggt ctg gtg cca GCC ggt agt act gca at) and *pET41 delete THRB rev* (at tgc agt act acc GGC tgg cac cag acc cg). The SUMO gene sequence was amplified from vectors containing the corresponding ESTs using the primers *SUMO1_NcoI_Start* (atatat ccatgg ga tta gtc cct agg gga agc atg tct gac cag gag gc) and *JO-211* (aaattt ggatcc atTA aac tgt tga atg acc ccc cg) for SUMO1; and *SUMO2_NcoI_Start* (tatat ccatgg ga tta gtc cct agg gga agc atg tcc gag gag aag ccc). The PCR reactions were performed using the following protocol:

Mix together
10 u Taq DNA polymerase
5 µL Buffer
1,6 µL each primer (from a 10pmol.µL^{-1} solution)
4 µL $_{d}$NTP (from a solution containing 2,5 mM of each nucleotide)
35,4 µL water
2 µL template (vector at 100 ng.µL^{-1})
The PCR program is: {[95°C, 1'], 26·{[95°C, 30''], [45°C, 45''], [72°C, 2']}, [72°C, 10'], [4°C, ∞] }

The PCR products and host pET41a vectors were digested with the restrictions enzymes *Nco*I and *BamH*I:

Add to 50 µL DNA to be digested (either the product of a PCR reaction or 100 ng.µL^{-1} vector solution)
10 u NcoI
20 u BamHI
6,5 µL BSA (from a 1 µg.µL^{-1} solution)
6,5 µL BamHI buffer from New England Biolabs

Incubate at 37°C for several hours

The digested vector was purified on a 1,2% agarose gel and extracted from it using QiaQuick (Qiagene) columns according to manufacturers instructions. The digested PCR products were purified using miniprep columns (Qiagene) according to manufacturer's instructions to a final volume of 50 µL. The PCR product and the vector were then ligated:

Mix
1,5 µL T4 DNA ligase
4,6 µL Ligase buffer
8µL purified digested PCR product
32 µL purified digested vector
Incubate overnight at 16°C

The ligation product was introduced in *E. coli* Rosetta cells by heat shock technique:

5µL of ligation reaction are added to an aliquote of chemically competent cells which are then placed on ice for 20-40 minutes, heated at 42°C for 45 seconds in a water bad and cooled on ice. 400 µL of LB medium without antibiotics are added to the cells which are agitated at 37°C for 1 hour and then plated on LB-Kan plates and grown overnight. Single colonies were picked for further culture.

The constructs were checked by sequencing (in house, using the BigDye terminator system from AbiPrism). The constructs were expressed in *E. coli Rosetta* cells on either LB or 2YT medium (for non marked protein), or minimal medium with $^{15}NH_4Cl$ and $^{13}C_6$-Glucose for production of ^{15}N and ^{15}N +^{13}C labeled protein. Cultures were grown in 3L flasks (~1.5 L culture per flask) at 37°C under agitation (100-120 rpm). Growth was monitored by following the optical density at 600 nm ("OD_{600}") of culture probes. When the culture reached OD_{600}~0.8, the culture were induced by addition of 1mM IPTG and further cultivated for 6-10 hours. Cells were harvested by centrifugation, resuspended in PBS buffer and sonicated. The cell debris were

pelleted by ultracentrifugation (17000 rpm in a SS-34 rotor) and the recombinant protein was purified from the supernatant on GSH resin (Amersham) according to the manufacturer´s instruction. The purified fusion protein was cleaved with thrombin (5 units per liter of cell culture) for 4 hours at room temperature, leaving the extension Gly-Ser in N-term of the full length SUMO, 30 and 5 KDa cutoff concentrators (Amicon) were used for tag removal and concentration. Purity was checked visually on SDS-PAGE gels. The absence of many unaccounted for peaks on the HSQC spectra was a further proof of the purity of the obtained protein. MALDI mass spectrometry was used to check the integrity of the obtained SUMO. The obtained protein was lyophilysed and stored at -20°C.

2. NMR Spectra acquisition and assignment

All measurements were made at 27°C in 25mM Phosphate buffer (Natrium salt) at pH 7. Triple resonance and 2D experiments were performed on a Varian Inova 600 equipped with shielded *Z* gradients. The water resonance was suppressed using the WATERGATE sequence or by presaturation. Prior to Fourier transformation all spectra were multiplied by a cosine bell function. Generally 2048x256 data points were used for acquisition of HSQC spectra and 1024x12x80 for three dimensional edited spectra. 3D NMR spectra were processed using the standard Bruker software XWINNMR. 2D NMR spectra were processed using NMRpipe (Delaglio *et al.*, 1995) (relevants parameters are listed in the fid.com and ft2.com scripts bellow). Analysis and visual representation of two-dimensional spectra were performed using Sparky (Goddard T.D. and Kneller D.G.) and three-dimensional spectra were analyzed with the program Aurelia (Bruker) on O2 and Octane workstations (Silicon Graphics Inc.). Assignment of SUMO2 was generated using the spectra HNCA, HNCACB, CBCA(CO)NH, HC(CO)NH, C(CO)NH.

Scripts:

Fid.com (transforms a spectrum in .fid format –the “Varian” format- to a spectrum in “nmrPipe format” which it stores in test.fid):

```
#!/bin/csh

var2pipe -in
#the following line give the location of the input file
```

```
/work/rabiller/nmr/sumo2/spectra/SUMO2_PPIAS_2004_11_18/SUMO2_and
_50uL_PPIAS_2004_11_18.fid/fid  \
  -xN              1024  -yN              256  \
  -xT               512  -yT              128  \
  -xMODE        Complex  -yMODE        Complex  \
  -xSW         9000.900  -ySW         2200.000  \
  -xOBS         599.835  -yOBS          60.787  \
  -xCAR           4.754  -yCAR         114.125  \
  -xLAB              H1  -yLAB            N15  \
  -ndim               2  -aq2D          States  \
  -out test.fid -verb -ov

sleep 5
```

ft2com (performs the processing of the input spectrum –here test.fid- and stores the processed spectrum in test.ft2)

```
#!/bin/csh -f

#
# Basic 2D Phase-Sensitive Processing:
#   Cosine-Bells are used in both dimensions.
#   Use of "ZF -auto" doubles size, then rounds to power of 2.
#   Use of "FT -auto" chooses correct Transform mode.
#   Imaginaries are deleted with "-di" in each dimension.
#   Phase corrections should be inserted by hand.

nmrPipe -in ./test.fid                  \
| nmrPipe  -fn SOL                              \
| nmrPipe  -fn POLY -time -verb                 \
| nmrPipe  -fn MAC -macro $NMRTXT/ranceY.M -noRd -noWr       \
| nmrPipe  -fn SP -off 0.5 -end 0.95 -pow 2 -c 0.5      \
| nmrPipe  -fn ZF -auto                            \
| nmrPipe  -fn ZF -zf 2                           \
```

```
| nmrPipe  -fn FT -auto                                  \
| nmrPipe  -fn PS -p0 -88 -p1 -25 -di -verb              \
| nmrPipe  -fn EXT -x1 6.2ppm -xn 11.3ppm -sw -verb      \
#| nmrPipe  -fn EXT -x1 6.1ppm -xn 10.8ppm -sw -verb     \
#| nmrPipe  -fn EXT -x1 -0.4ppm -xn 6.0ppm -sw -verb     \
| nmrPipe  -fn TP                                       \
#| nmrPipe  -fn LP -ps0-0 -verb                          \
| nmrPipe  -fn SP -off 0.5 -end 0.95 -pow 2 -c 0.5      \
| nmrPipe  -fn ZF -auto                                 \
| nmrPipe  -fn ZF -zf 2                                 \
#| nmrPipe  -fn ZF -size 32                              \
| nmrPipe  -fn FT -auto                                 \
| nmrPipe  -fn PS -p0 0 -p1 0 -di -verb                 \
#| nmrPipe  -fn REV                                      \
| nmrPipe  -fn TP                                       \
#| nmrPipe  -fn EXT -y1 5.4ppm -yn 10.5ppm -sw -verb     \
#| nmrPipe  -fn POLY -auto -verb                         \
#| nmrPipe  -fn POLY -auto -xn 4.72ppm -verb             \
#| nmrPipe  -fn POLY -auto -x1 4.62ppm -verb             \
  -ov -out ./test.ft2
```

3. NMR titration experiments

For peptide titration experiments HSQC spectra were performed on ^{15}N labeled SUMO1 or 2 (300 mM in 25 mM Phosphate buffer pH7). Unlabeled peptide (Thermo Electron GmbH) were titrated to the protein to reach a final Peptide:SUMO ratio of 1,36:1.

K_D values were measured by two different methods, depending on the exchange regime. For amino acids in fast exchange regime, at each titration step, the distance of each peak from its original position was measured using the normalization proposed by (Ayed *et al.*, 2001) The obtained curves were fitted to a Hill-4-parameters model using the software SigmaPlot. For amino acids in slow exchange regime, we estimated $k_e=1(k_{ex}/k_{on})$, where k_{ex} is the inverse of the lifetime of the

protein-peptide complex. In the case of slow exchange, the system is diffusion limited, giving k_{on}~10^8 M.s-^{1}, and $k_{ex}<\pi\Delta\upsilon/\sqrt{2}$, $\Delta\upsilon$ being the chemical shift difference in Hertz between the resonance without peptide and the resonance in saturated conditions for the considered atom. Therefore, k_e ~2.22 10^{-5} $\Delta\upsilon$ in the case of slow exchange.

4. Tryptic digestion of PIAS and detection of the resulting fragments by MALDI spectrometry

HEK 293T cells were transfected with Flag-PIAS, lysed and the Flag-PIAS expression was checked with Western blot analysis as described before (Haglund *et al*). Immunoprecipitation of Flag-PIAS was done with M2-agarose from Sigma according to manual instruction. Immunoprecipitated PIAS was loaded on SDS-PAGE gel and stained with Coomassie (this part of the work was performed by Christina Hecker).

The band containing PIAS was cut out of the gel and transferred into a microtube. As a negative control, a same sized bit of the gel cut from a region containing no protein and further handled as the probe. Except the digestion itself, all further steps were performed at room temperature under agitation (1000 rpm). The band was destained overnight using 0.5 mL of Acetic Acid/Methanol/Water 1:2:7 (V:V:V). It was washed for 4 hours with water, dried in Speed Vac, washed with (50% Acetonitrile 50% 0.2 M NH_4 HCO_3 in water, pH8.9) and dried in Speed Vac again. The band was soaked with 15 µL of 0.2 M NH_4 HCO_3 in water, pH 8.9 containing 33 micrograms.mL^{-1} trypsin proteomic grade (Sigma). The gel was reduced into little bits using the heat-rounded tip of a pasteur pipette and incubated at 37°C for 24 hours. The bits were washed with water for 2 hours, dried in speed vac, and covered with 50 microliters of formic Acid/Water/Isopropanol 1:3:2 (V:V:V) saturated with alpha-Cyano-4-hydroxycinnamic acid (CHCA). After overnight incubation, tubes were opened to allow crytallisation of the CHCA together with the extracted peptides. The obtained solution was pipetted onto the MALDI plate avoiding to pipette gel bits. Measurements were made in reflector positive mode, low mass gate set at 500 Da, and monitoring the 1 Kda-3KDa range.

Results

1. The SIM contains a phosphorylation site

Most of the SIM identified in the proteins obtained by Yeast Two Hybrids experiments contain several successive serines and occasional threonines between their hydrophobic core and negatively charged tract. As this looks typical of phosphorylation sites, the phosphorylation of SIM was investigated by bioinformatics and experimentally.

The program NetPhos [Blom *et al*, 1999; www.cbs.dtu.dk/services/NetPhos] was used on the different SIM, nearly all of which were predicted to contain at least one phosphorylation site with a probability higher than 90%. The recurrence of this prediction for all SIM makes it very unlikely that it is a false positive result. Phosphorylation can therefore be viewed as a constitutive characteristic of SIMs.

To experimentally verify the phosphorylation state of the SIM of PIAS, MALDI fingerprinting experiments were made on PIAS produced in mammalian cells (see Material & Methods section 4). Immunoprecipitated and SDS-PAGE purified PIAS given by Christina Hecker was trypsin digested in the gel and extracted with a water/acetic acid/ethanol mix. The particularity of this protocol is that the peptides are extracted from the gel directly by the MALDI-matrix solution, yielding much better result than the standard procedure [ref] used in the first place. Analysis of the MALDI spectra obtained for the extracts showed low intensity peaks at masses expected for the Tryptic peptides containing the SIM with and without phosphorylation (table 2). Such peaks could not be found in the negative control (empty gel piece handled the same way as the PIAS-containing gel bands). This shows that at least a fraction of PIAS is phosphorylated *in vivo*, but doesn't give any information on which of the three phosphorylable serines is actually phosphorylated.

2. Structure of the SUMO Interacting Motif

As described in the introduction, the SUMO interacting motif (SIM) appears to be limited to a peptide less than 25 amino acids in length within the proteins. This small size makes it doubtful that the SIM be a domain of its own right. One may ask whether the SIM is completed by further features in the three dimensional structures of the proteins containing it.

Models of the three dimensional structure of the proteins retrieved from the Yeast Two Hybrids experiments were produced using the program 3D-PSSM [Kelley *et al*, 1999].

				Well62_b_0003			Well62_b_0001			Well62_b_0006			Well62_b_0005		
Peptide sequence	Position	Unphosphorylated mass	Phosphorylated mass	m	m+1	m+2	m	m+1	m+2	m	m+1	m+2	m	m+1	m+2
KIDVIDLTIESSSDEEEDPPAKR	477-499		2666.238	2665.67	2666.74	2667.6				2665.4	2666.34	2667.27	2665.4	2666.34	2667.27
KIDVIDLTIESSSDEEEDPPAKR	477-499	2586.278		2585.65	2586.7	2587.74									
IDVIDLTIESSSDEEEDPPAKRK	478-500		2666.238												
IDVIDLTIESSSDEEEDPPAKRK	478-500	2586.278													
KKIDVIDLTIESSSDEEEDPPAK	476-498		2638.2319								2639.62	2640.58		2639.62	2640.58
KKIDVIDLTIESSSDEEEDPPAK	476-498	2558.2719													
IDVIDLTIESSSDEEEDPPAKR	478-499		2538.1431				2539.53	2540.56	2541.5						
IDVIDLTIESSSDEEEDPPAKR	478-499	2458.1831								2457.78	2458.78	2459.73	2457.78	2458.78	2459.73
KIDVIDLTIESSSDEEEDPPAK	477-498		2510.1369												
KIDVIDLTIESSSDEEEDPPAK	477-498	2430.1769													
IDVIDLTIESSSDEEEDPPAK	478-498		2382.042										2381.76	2382.73	2383.79
IDVIDLTIESSSDEEEDPPAK	478-498	2302.082													

				Well62_b_0008			Well62_b_0009				Well62_b_0010	
Peptide sequence	Position	Unphosphorylated mass	Phosphorylated mass	m	m+1	m+2	m	m+1	m+2	m	m+1	m+2
KIDVIDLTIESSSDEEEDPPAKR	477-499		2666.238							2666.44	2667.45	2668.4
KIDVIDLTIESSSDEEEDPPAKR	477-499	2586.278					2585.45	2586.45				
IDVIDLTIESSSDEEEDPPAKRK	478-500		2666.238									
IDVIDLTIESSSDEEEDPPAKRK	478-500	2586.278										
KKIDVIDLTIESSSDEEEDPPAK	476-498		2638.2319									
KKIDVIDLTIESSSDEEEDPPAK	476-498	2558.2719										
IDVIDLTIESSSDEEEDPPAKR	478-499		2538.1431							2538.61	2539.65	
IDVIDLTIESSSDEEEDPPAKR	478-499	2458.1831			2458.82	2459.77						
KIDVIDLTIESSSDEEEDPPAK	477-498		2510.1369									
KIDVIDLTIESSSDEEEDPPAK	477-498	2430.1769										
IDVIDLTIESSSDEEEDPPAK	478-498		2382.042	2382.72	2383.74	2384.65	2382.91	2383.87		2382.39	2383.37	
IDVIDLTIESSSDEEEDPPAK	478-498	2302.082										

Table 2: Peptides corresponding to the SIM of PIAS detected by MALDI fingerprinting on immunoprecipitated PIAS from mammalian cells (HEK 293T). The position of these peptides in PIAS sequence is given in the second column, their expected mass with and without phosphorylation in the third and fourth column. Since the signal was very weak, only masses with which two isotopic masses could be associated were retained. The mass measured for the peptide with no, one or two isotopes are refered to as m, m+1 and m+2. The annotations “well62_b_003” refer to the well on the MALDI plat in which the measurement was made, and the serial number of the measurement.

No constant three dimensional elements could be distinguished in those models. In all cases, the SIM is found in unstructured regions outside the core of those proteins. The core of the SIM and the amino acids surrounding it have a composition characteristic of β-strand but are often in a position that makes it unlikely for them to be part in a β-sheet of the protein. The rest of the SIM can be in all cases predicted to adopt random coil conformation. The SIM is therefore likely to be the sole element responsible for SUMO binding. Its short size and lack of tertiary structure make it a perfect candidate for peptide binding studies: instead of investigating the binding of a whole protein to SUMO, which brings about all difficulties associated with high concentrations of a protein in solution, one can observe the binding of a short peptide derived from its SIM to SUMO and obtain relevant results.

3. Design of PIAS- and TTRAP derived peptides

In this study, peptides derived from the SIM of PIAS (SwissProt entry O75928) and TTRAP (SwissProt entry O95551) were used. The sequences and features of

Fig. 20: Peptides used in this study. The hydrophobic core is highlighted in yellow. Positively charged amino acids are in blue, amino acids bearing a negative charge in red. Phosphate groups are symbolized by a circled “P” above the amino acids bearing them.

those peptides are presented on fig. 20. PIAS_long contains all of the SIM of PIAS. PPIAS has the same sequence as PIAS_long and is phosphorylated on serine 476, which is the most probably phosphorylated serine in PIAS. PIAS_short contains the hydrophobic core of the SIM and a few amino acids around it, but lacks the negatively charged tract. TTRAP derived peptides were designed referring to the

PIAS derived peptides. The four amino acids making the hydrophobic core of TTRAP's SIM are the same as in PIAS, but in inverted order. In absence of a clear negatively charged tract to give an orientation the TTRAP_long peptide was therefore designed as the hydrophobic core of the SIM and enough amino acids in N-term of it to have the same length as PIAS_long (in PIAS_long, the rest of the amino acids are in C-term of the hydrophobic core). Two further TTRAP derived peptides were designed to be comparable to PIAS_short: TTRAP_short_C-side and TTRAP_short_N-side. Both are the same length as TTRAP and contain the hydrophobic core plus a few amino acids in C-term of it (like in PIAS_short) or in N-term of it (orientation oppose to that of PIAS_short). Since no phosphorylation site can be found in TTRAP's SIM, no phosphorylated TTRAP peptide was designed. All those peptides were ordered in highly purified form from the company Thermo Electron GmBH.

4. Cloning and protein production

This study required large quantities of variously isotope-marked full length SUMO1 and SUMO2. It was therefore chosen to produce the needed SUMO as recombinant protein in the *Escherichia Coli* bacteria optimized for protein production. To simplify the purification process and guarantee high purity and yield, it was chosen to produce SUMO in a GST fused form. To reduce the complexity of the NMR spectra as much as possible and avoid interferences of the GST tag, this tag had to be removed leaving as few exogenous amino acids as possible on SUMO. In prevision of further studies, SUMO3 and –4 were cloned as well. The experimental procedures followed for cloning and protein purification are described in the section 1 of the Materials & Methods part.

The pET41a vector from Novagen was selected for cloning of SUMO. It contains an endogenous Thrombin cutting site to be used for tag removal. However, this site is several amino acids away from the closest recombinant protein insertion site. It was therefore mutated to prevent thrombin cleavage and a new thrombin cutting site was introduced immediately in N-term of SUMO (fig. 21) in the cloning process.

ESTs containing the mRNA sequence of SUMO1, -2 and -3 were purchased from Invitrogene. The SUMO sequences to be cloned were amplified by PCR from those ESTs using primers bearing restriction sites (for cloning into the vector) and, in

the case of the forward primer, the sequence of a thrombin cutting site. The correctness of the SUMO sequence and their insertion in the vector was checked by sequencing. SUMO1, -2, -3 clones were produced.

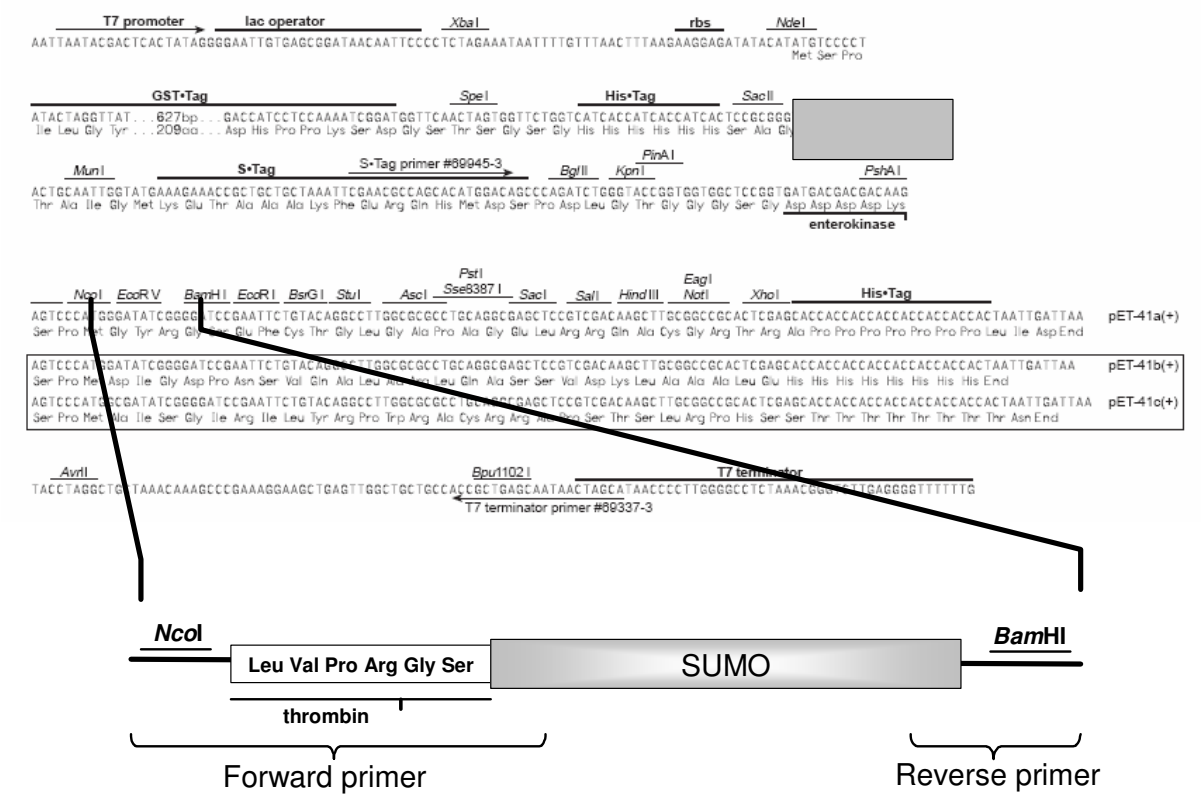


Fig. 21: Cloning of SUMO into the multiple cloning site of the pET41a vector. The original thrombin cutting site coded by the vector (boxed) is first mutated to prevent thrombin digestion at this site in the recombinant protein. A synthetic DNA construct coding for a thrombin cutting site directly followed by full-length SUMO between a NcoI and a BamHI restriction sites is inserted in the pET41a vector. The extent of the primers used for it synthesis it is represented under the scheme the inserted construct.

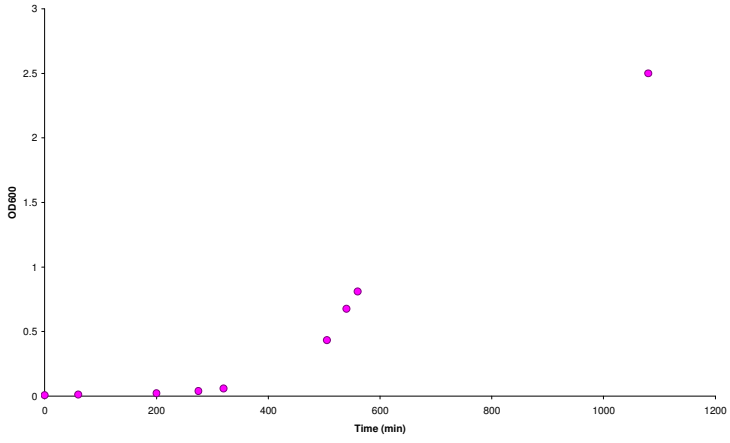
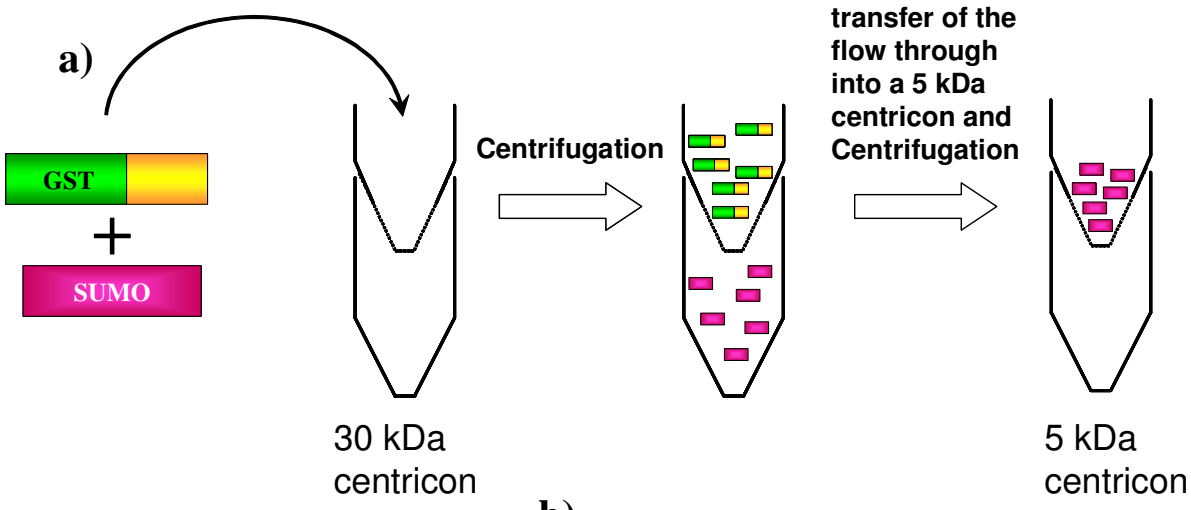


Fig. 22: Typical evolution of a Rosetta cell culture used for SUMO production. In this particular case, The cells produce SUMO1, and are grown on 15N labeled minimal medium. The culture was induced after 560 minutes by addition of 1mM IPTG. The last point (1080 min) is the measurement made just before harvest. A sample of the culture was reserved before induction and cultivated without IPTG. At harvest time, its OD600 was 3.4. OD values higher than 1 were calculated from values measured on diluted culture samples.

A different route was chosen for cloning SUMO4. No EST was available for this gene, but fortunately SUMO4's gene contains no exons, meaning that its translated sequence can be obtained by amplifying genomic DNA without any further processing. To this purpose, a commercial kit (Sigma) was used to extract

genomic DNA from the experimentator's hair roots. SUMO4 gene sequence was amplified from this genomic DNA by PCR and cloned in the same way as the other three SUMO isoforms. Sequencing showed that no clone without nonsense mutation was obtained. At that point, it turned out that it is doubtful that SUMO4 is a genuine gene, and no further work on SUMO4 cloning was made.

Rosetta cells (Novagen) were transformed with the obtained pET41a vectors containing SUMO1 and SUMO2 and grown on 2YT medium (for obtention of plain protein) or on minimal medium with appropriately labeled Glucose and NH_4 as sole source of Carbon and Nitrogen respectively (for obtention of ^{15}N and/or ^{13}C labeled protein). The bacteria were allowed to grow until they reached an optical density at 600nm of around 0.8. At that point, 1 mM IPTG was added to the culture to induce the bacteria to produce the recombinant protein, and the bacteria were further incubated for 6-10 hours (fig. 22), harvested by centrifugation and lysed by sonication. The SUMO-GST fusion protein was then purified on a GSH column, eluted and treated with thrombin (alternatively, the thrombin cut was performed on



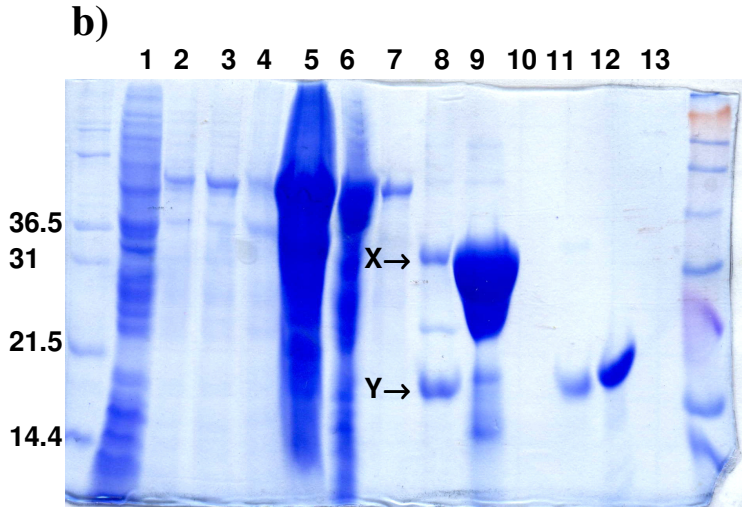


Fig. 23: a) Use of centricons to separate SUMO from the GST-Tag and concentrate it after purification and Throbin cutting of the fusion protein. b) Coomasie gel with probes taken at different steps of SUMO1 production. Whole bacteria before induction (1), at harvest without (2) and with (3) induction, Pellet (4) and supernatant from centrifugation of the cell lysate. Flow Through (6), wash (7), thrombin cut (8), elute (9) and clean (10) fraction from the GSH purification (the thrombin cut was made while the protein was still bound to the resin, the cut fraction contains the collected SUMO). The fractions retained by the 10 kDa and 5 kDa centricons are in lane 11 and 12 respectively. Band X contains the fusion protein, Y contains SUMO

the column). Tag removal, concentration and setting in measurement buffer conditions were performed in a two-steps filtration (fig. 23a). In the first step, the thrombin-treated elute was passed through a 30kDa cutoff concentrator (”centricon”). This filter size is large enough to allow the GST-Tag free SUMO to go through, and small enough to retain GST and GST-SUMO fusion protein as well as thrombin and large molecular weight impurities. In the second step, the flow through of the first step was passed through a 10 kDa concentrator. This filter size is small enough to retain SUMO, but not the GSH. This allows concentration of SUMO and, by repeated cycles of dilution in the measurement buffer and subsequent concentration, to set SUMO in the measurement conditions without having to perform any dialysis step. Coomassie-stained SDS page gels show with probes taken at the different steps of the protein production and purification show that SUMO pure as far as it can be determined from this method (i.e. purity is higher than 90%).

5. Assignment of the resonances of SUMO2 atoms

To investigate the effect of peptide binding on specific atoms of a protein by NMR, one needs to know the resonance frequencies of each of those atoms. For this purpose, a set of 3D NMR spectra of $^{15}N/^{13}C$ labeled SUMO2 was recorded. (see M&M 2.) The assignment was based on the HNCACB spectrum, following the chain tracing procedure described in the introduction. The program PaarFind (Karl-Heinz Müller, not published), which is able to pick peaks and identify the most obvious sequences of peaks quartets was used for a first approach of the HNCACB spectrum, its output being manually checked, assigned and completed. In order to calculate a high resolution structure of SUMO2, the resonances of a number of the backbone and the side chain atoms were assigned using the other 3D spectra. The publication of a crystal structure of human SUMO2 [Huang *et al*, 2004] made the structure calculation superfluous, and no further step was taken in this direction. The assigned HSQC spectrum of SUMO2 is shown in fig. 24. As several assignments of SUMO1 have already been published [Jin *et al*., 2001; mac Auley *et al*., 2004, by far the most helpful being the one published by mac Auley *et al*], it was not necessary to produce a new one.

In general, atoms do not have exactly the same resonance frequencies in different spectra. Those differences are generally small, but can cause some

difficulties in assigning the resonances in one spectrum using the frequencies found in another one. To simplify the assignment of HSQC spectra using values from the HNCACB spectrum or obtained from literature, a graphic representation of the ^{1}H and

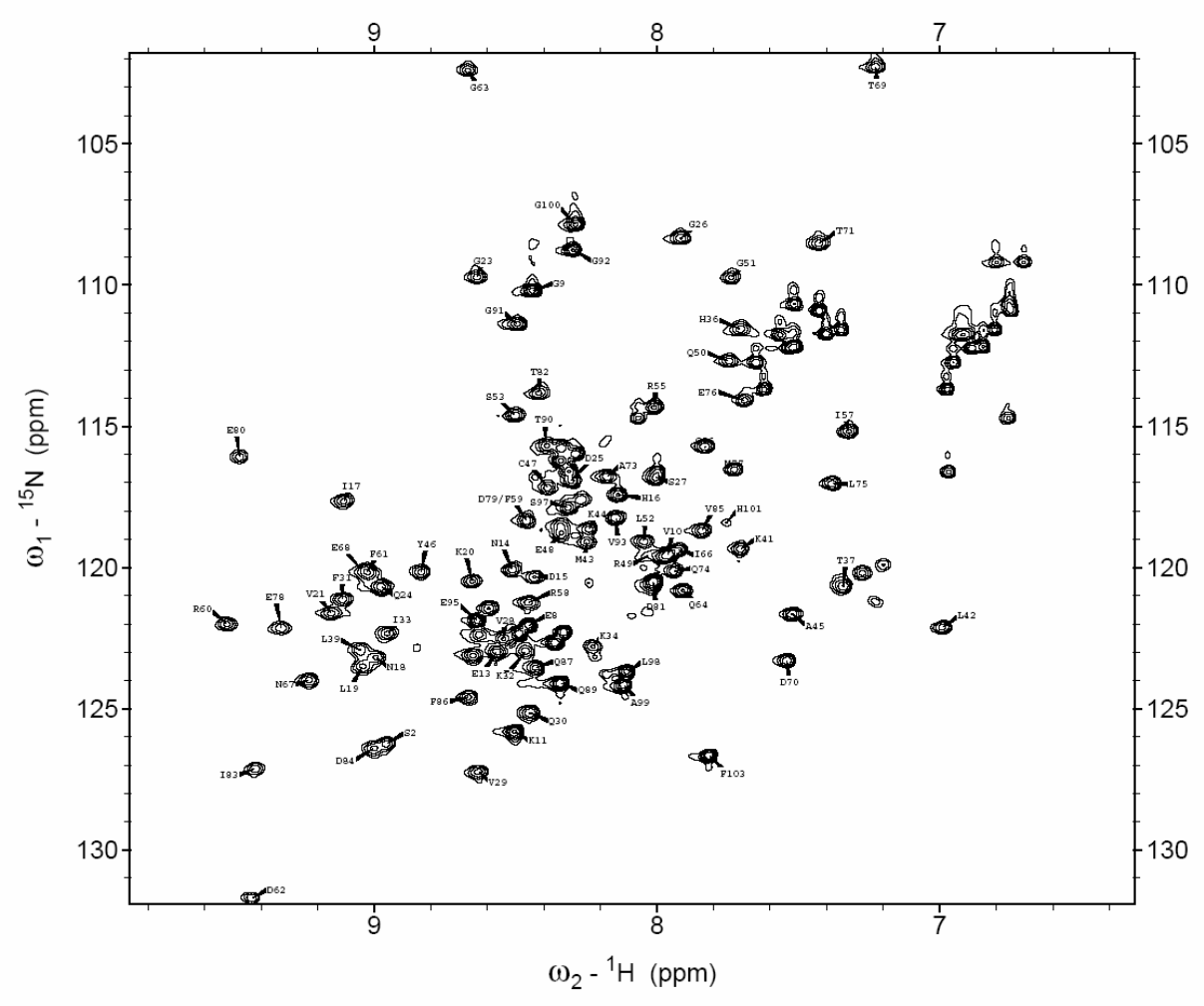


Fig. 24: Assigned HSQC spectrum of SUMO2.

^{15}N resonances obtained from the 3D spectra spectrum was produced with the program SigmaPlot and overlaid on the actual HSQC spectra. This allows seeing at one glance all expected and measured peak, which is of great help for sorting out the assignment ambiguities.

6. Measuring the effect of the binding of a peptide on all amino acids of a protein

As exposed in the introduction, the binding of a ligand modifies the electromagnetic environment of the atoms of a protein, especially of those in the neighborhood of the binding site. This results in modifications of the resonance of those atoms in NMR. Those modifications can be observed by taking a series of

NMR spectra in the course of the titration and observing the evolution of the resonances of interest. This method, originally -and in most cases incorrectly- known as “SAR by NMR” has become very popular since its publication by [Shuker *et al*, 1996] and is used with HSQC spectra which have the advantage of being quickly measured and reflect the behavior of the backbone of the protein.

For the peptide titration experiments used in this study, a 1.35-fold molar excess of peptide was added in 10 steps to a 0.3mM SUMO solution and a HSQC spectrum was recorded after each addition. In the overlay of those spectra, peaks are observed to have different behaviors. Some are constant, other gradually change position, disappear and then reappear at a new position, or simply disappear (fig. 25). This section will expose how this can be interpreted and how those spectra are analyzed to gain quantitative information on binding.

The obviously simplest –and least informative- case is that of peaks that

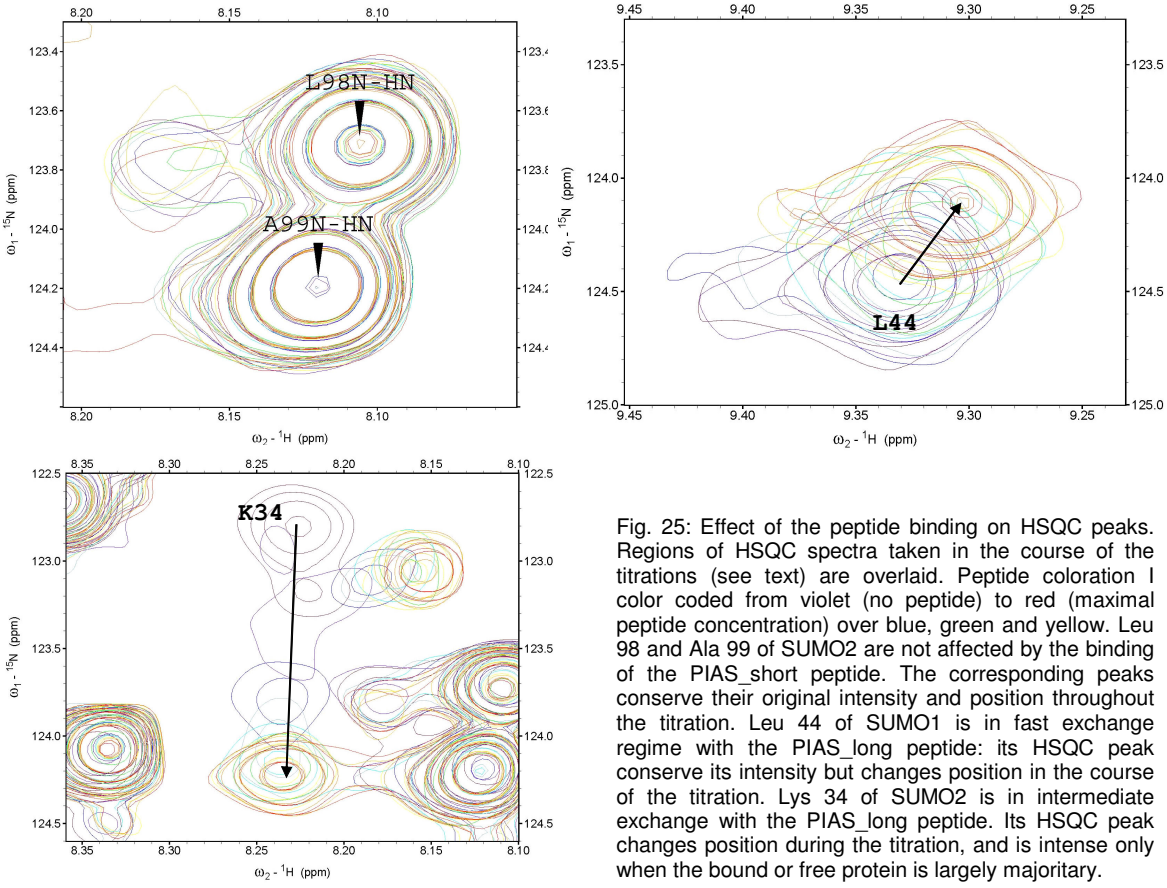


Fig. 25: Effect of the peptide binding on HSQC peaks. Regions of HSQC spectra taken in the course of the titrations (see text) are overlaid. Peptide coloration I color coded from violet (no peptide) to red (maximal peptide concentration) over blue, green and yellow. Leu 98 and Ala 99 of SUMO2 are not affected by the binding of the PIAS_short peptide. The corresponding peaks conserve their original intensity and position throughout the titration. Leu 44 of SUMO1 is in fast exchange regime with the PIAS_long peptide: its HSQC peak conserve its intensity but changes position in the course of the titration. Lys 34 of SUMO2 is in intermediate exchange with the PIAS_long peptide. Its HSQC peak changes position during the titration, and is intense only when the bound or free protein is largely majoritary.

change neither in position nor intensity in the course of the titration. This visualizes that the resonances of the corresponding atoms stay constant in frequency and intensity, meaning that the environment of those atoms is not modified by the binding.

In all other cases, the changes in position and/or intensity of HSQC peaks show that the corresponding atoms are affected by the binding. This is the result of the atoms changing environment ("free protein" and "bound protein") at different rates ("exchange rates") when the ligand is added. This phenomenon is called "chemical exchange". It gives access to information on the exchange rate (the frequency at which the spin changes environment) and therefore gives a possibility to calculate exchange constants (k_e) for individual atoms in a protein. Exposing how this can be done will require a preliminary explanation on chemical exchange rates and their effects on peak shape and position.

The prerequisite for observing an effect of chemical exchange on NMR resonance frequency and intensity is that the resonance frequency is in the order of magnitude of the exchange rate. This condition fulfilled, let us consider a spin moving between two different environments (it may be, for example, found in the side chain

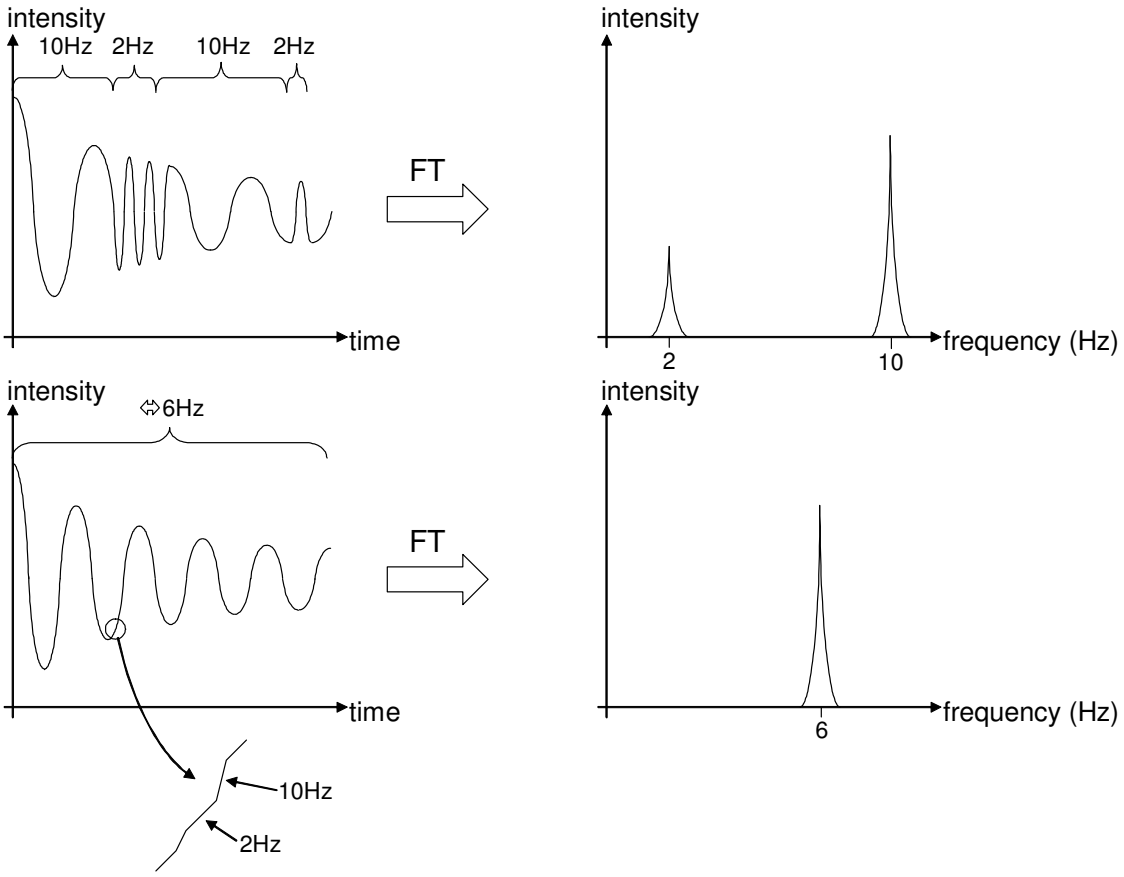


Fig. 26: Resonances of spins changing environment at different rate. The upper part of the figure shows a Spin changing environment at a rate slower than the difference of frequency between the resonances in the two environments. As seen on the resulting FID on the left, the Spin has time to make several revolutions in each of the environments. Fourier transformation of this FID (upper right) gives two peaks, one for each of the resonance frequencies, weighted by the time spent by the Spin in each environment. The lower part of the figure shows the case of a Spin changing environment at a frequency much higher than the difference between its resonances frequencies in the two environments. In this case, the Spin has time to make only a small fraction of revolution in each environment, resulting in a FID that appears to have an average frequency, weighted by the time spent by the Spin in its two environments. Fourier ransformation of this FID (lower right) gives a single peak at the average frequency.

of an amino acid moving between a buried and a solvent-exposed conformation). In one of those environments, the spin will have a resonance frequency of 10 Hz, in the other one of 2Hz, the average lifetime of each state being 2s. This lifetime corresponds to an exchange rate of ½ s^{-1}=0.5 Hz, which is significantly smaller than the difference between the resonance frequencies of the spin in its two environments (10-2=8Hz). Such a situation is called “slow exchange”. The FID curve obtained when making an NMR measurement on such a spin will be composed of segments with a frequency of 10 Hz and segments with a frequency of 2 Hz (fig. 26). The FID for a number of such spins would contain both frequencies -2 and 10 Hz- which would, by Fourier transformation, give two peaks, one at 2 Hz, the other at 10 Hz (fig. 26). If for whatever reason (they have different lifetime, one environment is encountered more often than the other…) the two states are differently populated, the intensity of the corresponding peaks will be accordingly unequal.

We have seen that if the exchange rate between the two environments of a spins is much smaller than the difference between its resonance frequencies in those two environments, one peak is observed at each of those frequencies in a NMR spectrum. What happens if the exchange rate is much higher than this frequency difference, a situation called “fast exchange”? In this case, the spin will be able to change many times of environment in the time it needs to complete a single revolution in either of those environments. This will result in a FID for a single spin made of very short segments of FIDs of both frequencies, which looks like a FID containing a single frequency –the average of the resonance frequencies in the two environments, in the present case (2+10)/2=6 Hz. By Fourier transformation of this FID, a single peak at 6 Hz (fig. 26) will be obtained. As before, this average is weighted if the to environments are unequally populated.

If, finally, the exchange rate becomes of the same order than the difference between the resonance frequencies of a spin in its two environments –a situation called “intermediate exchange”, the corresponding NMR peaks will broaden and their position will move toward the average between the resonance frequencies in the two environments. The origin of this phenomenon is the dephasing arising at each transition between two resonance frequencies: each time a spin changes resonance frequency, it loses its original phase. In a population of spins, this results in a quickly decaying FID giving a broad NMR peak by Fourier transformation. It can be demonstrated that the maximum broadening is reached when the exchange rate is

equal to $\pi \cdot \Delta\upsilon/\sqrt{2}$, $\Delta\upsilon$ being the difference between the resonance frequencies in the two environments. At this point, the NMR peaks corresponding to the two environments merge into a single one at the average between the resonances frequencies in those two environments. Fig. 27 shows an example of evolution of the NMR spectra obtained when varying the exchange constant between two environments of a group of identical spins.

How does this relate to the behavior of NMR peaks in HSQC titrations?

The peaks changing position but not intensity clearly correspond to spins in fast exchange as described above. At the beginning of the titration, no peptide has been added, and all the spins in a particular amide group are in "free protein" environment. As peptide is being added, more and more of it becomes bound to the protein, and accordingly more and more of the considered spins find themselves in a "bound protein" environment until saturation is reached and the bound environment prevails. At each step of the titration, the HSQC peak produced by those spins occupy a position corresponding to the average of the resonances frequencies in the "free protein environment" and the "bound protein environment" states, weighted by the respective populations of those states (fig. 27). The exchange constant (k_e) associated with this behavior can be estimated as follows. Let P_f, P_b and P_t be the fractional concentration of free, bound ant total protein respectively. The peptide (L) binds to the protein (P) according to the simple balance equation

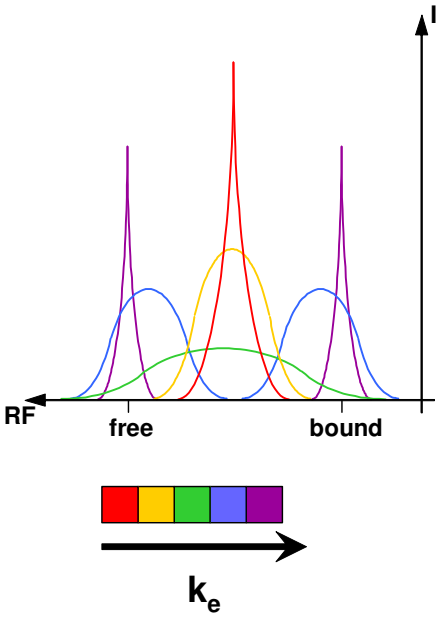


Fig. 27: Effect of the exchange constant k_e on the NMR peaks produced by a group of Spins exchanging between two environments (here bound and free protein).

$$P+L \rightarrow PL.$$

The constant associated with this binding is

$$k_e = [P]\cdot[L]/[PL]$$

Since

$$[L] = [L]_{initial} - [PL]$$

the expression of k_e becomes

$$k_e = [P] \cdot ([L]_{initial} - [PL]) / [PL]$$

which can be rearranged into ($[L_i]$ being the initial peptide concentration and [P]i the initial protein concentration):

$$k_e \cdot [PL] = [P] \cdot ([L]_i - [PL])$$

$[L_i]$ and $[P]_i$ are known –they are the concentration of peptide and protein introduced in the probe- and the position of the HSQC peak gives the rest of the information required to calculate k_e. The position of a fast exchange peak in the course of the titration is the weighted average between its position in the absence of ligand and its position at saturation, when all protein is bound. Let Δ_{max} be the distance between the HSQC peak in free protein and in bound protein, and Δ_{if} be the distance between the HSQC peak at an intermediate position and the HSQC peak in the free protein (fig. 28, the normalization proposed by [Mulder *et al*, 1999] was used). We have

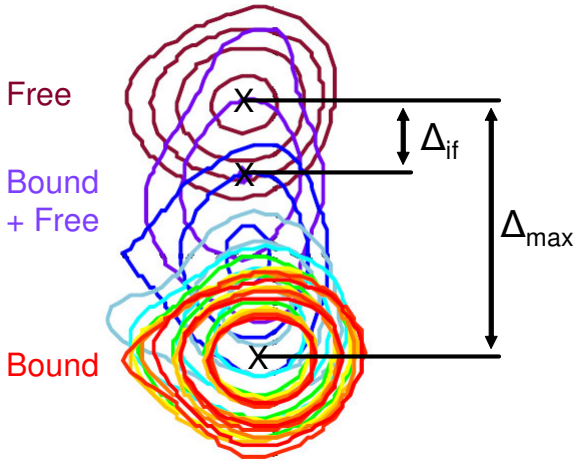


Fig. 28: Overlay of the HSQC spectra taken in titration of SUMO2 with PPIAS. Only the peak corresponding to Lys41 is shown. Δ_{max} (distance between the bound and free protein peaks) is constant. Δ_{if} (distance between the peaks for free protein and mixed bound and free protein) is different for each step of the titration. Here, Δ_{if} after the first peptide addition (mauve peak) is shown.

$$\Delta_{if} / \Delta_{max} = [PL]/([P]+[PL]) = [PL]/[P]_i = y$$

coming back to the expression of $k_e \cdot [PL]$ and dividing both sides by [P]i:

$$k_e \cdot [PL] / [P]_i = [P] \cdot ([L]_i - [PL]) / [P]_i$$

Replacing $[PL]/[P]_i$ by x, this becomes

$$k_e \cdot y = [P] \cdot ([L]_i / [P]_i - y)$$

There is still a stray [P] –the actual free protein concentration- which can be replaced by [P]i - [PL], yielding

$$k_e \cdot y = ([P]_i - [PL],) \cdot ([L]_i / [P]_i - y)$$
$$(k_e \cdot y) / [P]_i = (1 - y) \cdot [L]_i / [P]_i - y)$$

This develops into

$$y^2 - y \cdot (k_e + [L]_i + [P]_i) / [P]i + [L]_i / [P]_i = 0$$

the solution of this quadratic equation is

$$y = ((k_e + [L]_i + [P]_i) - \sqrt{((k_e + [L]_i + [P]_i)^2 - 4 \cdot [L]_i \cdot [P]_i)}) / (2 \cdot [P]_i)$$

Since Δ_{max} is constant for each amino acid, $y = \Delta_{if} / \Delta_{max}$ is always proportional to Δ_{if}. Hence, by plotting Δ_{if} against the total peptide concentration and fitting those points the equation above, the value of K_e can be calculated. However, in the present work, some titration curves did not follow expected the shape observed in SUMO2 titrations (fig. 29). The Hill-4-parameters equation was therefore used to fit all data. This equation contains more parameters than the equation presented above, but is equal to it when those parameters are equal to zero. It is therefore perfectly suited for fitting titration curves that have the shape expected for a A+B→AB binding, and proved to fit the other data with very good correlation factors.

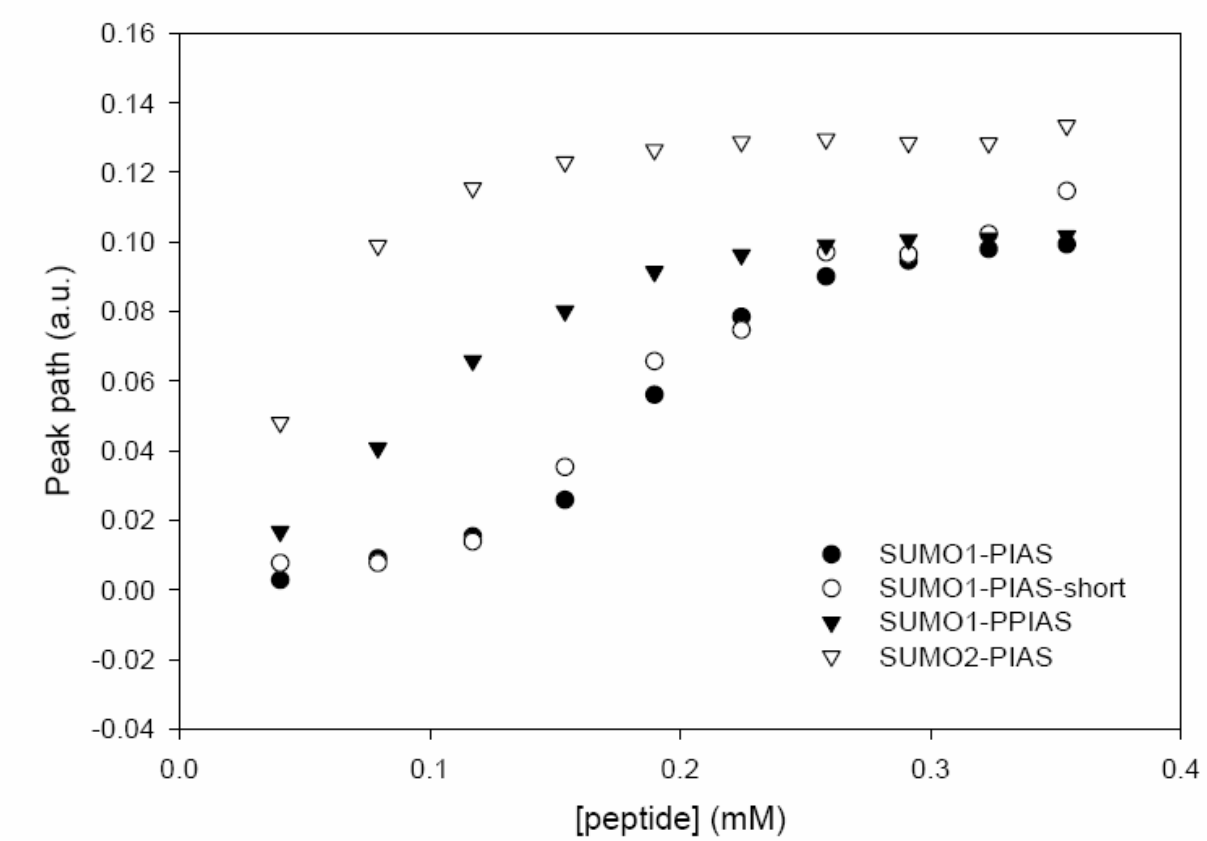


Fig. 29: Comparison of binding curves typical of different titrations. Values obtained with Glu49 of SUMO1 and Met43 of SUMO2 are plotted.

The case of peaks disappearing from one position to reappear at another position later in the titration is more delicate. Two possibilities can be envisaged: slow exchange or intermediate exchange. Let us examine what kind of peak behavior can be expected in either case.

At the beginning of a titration with a ligand causing the observed spins to be in slow exchange, there is very little ligand present, and most of the protein is in the free state. consequently, the NMR peak corresponding to the free state is very intense whereas the one corresponding to the bound state is very small. This situation changes in the course of the titration. More and more ligand becoming available, the fraction of bound protein increases and the fraction of free protein decreases accordingly, which result in increasing intensity for the peak corresponding to bound state and decreasing intensity for the peak corresponding to free state (fig. 30). This looks like the behavior of the “disappearing-reappearing peaks”, with two significant differences: slow exchange peaks would not change position, and at least at some point of the titration, both should be visible simultaneously.

At the beginning of a titration with a ligand causing the observed spins to be in intermediate exchange, there is very little ligand present, most of the protein is in the free state and a single peak is observed at the corresponding position. When more ligand is added, more and more of the spins have the opportunity to make incursions into the “bound protein” environment, which causes them to be out of phase with the rest of the spins. In consequence, the peak broadens and adopts a position intermediate between the “free protein” and “bound protein” positions. When the titration proceeds toward saturation, the “bound protein” environment becomes the most frequently occupied one, and incursions into the “free protein” environment become rarer. In consequence, the peak becomes sharper and its position becomes close to the “bound protein” position (fig. 30). This corresponds to the behavior of the experimental “disappearing-reappearing peaks”: the disappearance is caused by broadening which make the peak so broad and shallow that it is not detectable

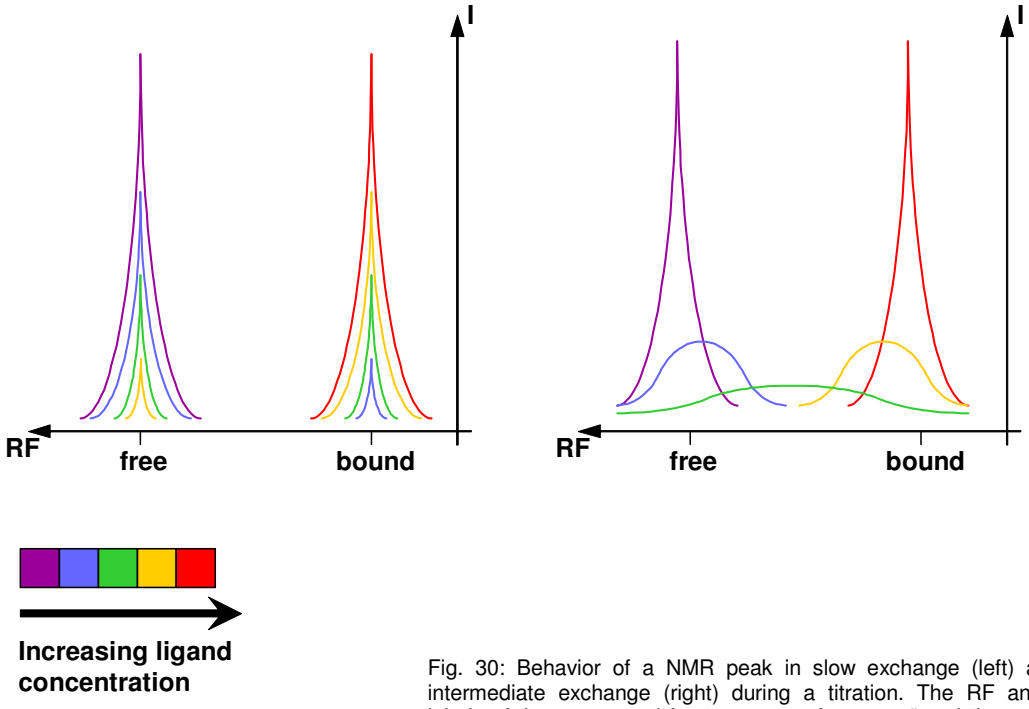


Fig. 30: Behavior of a NMR peak in slow exchange (left) and intermediate exchange (right) during a titration. The RF and I labels of the axes stand for “resonance frequency” and “intensity” respectively.

anymore. It can therefore be concluded that those peaks correspond to atoms in intermediate exchange. The value of the exchange rate (i.e. the rate at which the protein releases a bound ligand) Kex can therefore be estimated to be approximately equal to $\pi \cdot \Delta\upsilon/\sqrt{2}$ where $\Delta\upsilon$ is the difference of resonance frequencies between the free and bound protein states. Furthermore, it can be assumed that the peptide-

protein binding is limited by the diffusion rate. Kon, the rate at which the peptide contacts "encounter" the protein, is therefore of the order of 10^8 M.s-^{1}. Having done those assumptions on Kex and Kon, it is possible to estimate the value of the exchange constant k_e:

$$k_e = Kex/Kon = \pi \cdot \Delta\upsilon/\sqrt{2} / 10^8 = 2.22 \cdot 10^{-5} \cdot \Delta\upsilon$$

This was used to estimate the k_e of all atoms giving a peak that disappear in the course of a HSQC titration.

7. Effect of binding to PIAS- and TTRAP derived peptides on SUMO1 and SUMO2

For the sake of comparability, all titration experiments in this study were conducted in the same way. A 1.35-fold molar excess of peptide was added in 10 steps to a 0.3mM SUMO solution and a HSQC spectrum was recorded after each addition. The program Sparky [Goddard and Kneller] was used to overlay the spectra in order to visualize the behavior of the peaks in the course of the titration, and to determine the peaks position in each of those spectra. The peaks were assigned in each spectrum whenever possible (it happens regularly that two peaks adopt overlapping positions in the course of a titration, making it impossible to accurately determine the position of each).

In each titration, the initial spectrum, measured without peptide, was taken as a reference, and the distance of each peak in each of the spectra measured in the course of the titration to this original position (all in ppm) was measured using the normalization proposed by [Mulder *et al*, 1999]:

$$\Delta\delta = \sqrt{(\Delta\delta_{HN}^2 + (0.154 \cdot \Delta\delta_N)^2)}$$

where 0.154 is the scaling factor. The rationale for this normalization is the following. It is easily seen in any HSQC titration that upon ligand binding most peaks shift by a much greater number of ppm along the ^{15}N axis than along the ^{1}H axis. Part of the reason for this is the difference in the scale of the axes: 1ppm represents about 60 MHz on the ^{15}N axis and about 540 MHz along the ^{1}H axis. Hence, if ligand binding or whatever environmental condition had the same effect on N and HN resonance

frequencies, a scaling factor of 60/540 ≈ 0.113 would have to be used when combining N and HN shifts expressed in ppm. Yet there is *a priori* no reason why the resonance frequencies of N and HN atoms should be affected exactly to the same extent by their environment, which would result in the average frequency shift in Hz along the ^{15}N axis to be the same as the average frequency shift in Hz along the ^{1}H axis. To have an accurate estimation of the relative effects of environment on frequency shifts of N and HN atoms, Mulder *et al.* calculated the variance of the chemical shifts of al N and HN reported in the BMRB database, and found the variance of N resonnance frequencies (in ppm) to be 0.154 time the variance of the HN resonance frequencies. Therefore 0.154 is used as a scaling factor to ensure that the effect of binding on both resonances is equally taken into account.

The exchange constants k_e were calculated whenever possible for the single amino acids of SUMO in each titration using the methods exposed in the previous section. They are found to be in the low micromolar range for amino acids in intermediate exchange and in around 150 micromolar range for the vast majority of amino acids in fast exchange. For different reasons, the k_e value couldn't be estimated for some amino acids:

- the peak for this amino acid could not be identified in the HSQC spectrum
- the peak for this amino acid overlapped another peak making it impossible to determine its position
- the peak for this amino acid changed position so little that fitting couldn't be performed properly.
- The amino acid was too far from saturation, making the calculation of k_e imprecise.

The obtained k_e and Δ_{max} values are represented in the tables 3 and 4 (for SUMO1 and -2 respectively) and in graphic form (fig. 31 and 33) at the end of this section. Both representation of the results are linear, and do not take the three dimensional structure of SUMO into account. Yet a peptide binding site may consist of amino acids scattered along the sequence of SUMO but grouped together in its 3-dimensional structure. This called for a representation of k_e on the three dimensional structure of SUMO. Of course, writing its k_e value next to every amino acid would be utterly illegible, and adding graphical elements to symbolize it would make the structure of SUMO unrecognizable. It was therefore chosen to represent k_e values by a color coding. Using Yasara offers many advantages, but also imposes some

constraint on coloring possibilities: Yasara codes colors as an angle in degree corresponding to their position on a color wheel. It was therefore chosen to represent the range of k_e values as a color range going from icy blue for high values of k_e (corresponding to limited effect of peptide binding) to orange for low values of k_e (corresponding to strong effect of peptide binding), over blue, violet, pink and red. The k_e values calculated for amino acids in intermediate exchange being both imprecise and much smaller than those calculated for amino acids in fast exchange, they were all represented in yellow-green. k_e values that could not be determined were color coded by grey. A procedure for defining the color coding for each amino acid in fast exchange had then to be defined. It was important that the values clearly standing out were represented by colors clearly standing out, and with the gradation seen in the graphical representation. This proved impossible to achieve using a simple linear gradient of color: in such a gradient, if the very lowest values of k_e are represented in orange, all other values are represented in indistinguishable shades of another color. Therefore, a “square root gradient” was chosen, allowing the various low values of K_e to be clearly distinguished while the small variations of k_e in the bulk of SUMO are not exaggerated.

$$\text{Color code} = \text{int}\,[410 + (Z\ /\ |Z|) \cdot \sqrt{|Z|}]$$

where

$$Z = F \cdot (k_e - k_{e,\,ref})$$

F being a factor which value is chosen so that the obtained color range has the whished span. It must be clear that the color codes are specific to each titration, and a color does therefore not code for the same k_e values in two different titrations. Structures of SUMO1 (PDB code 1A5R) and SUMO2 (PDB code 1WM2) colored according to this method are presented below (fig. 32 and 34 respectively).

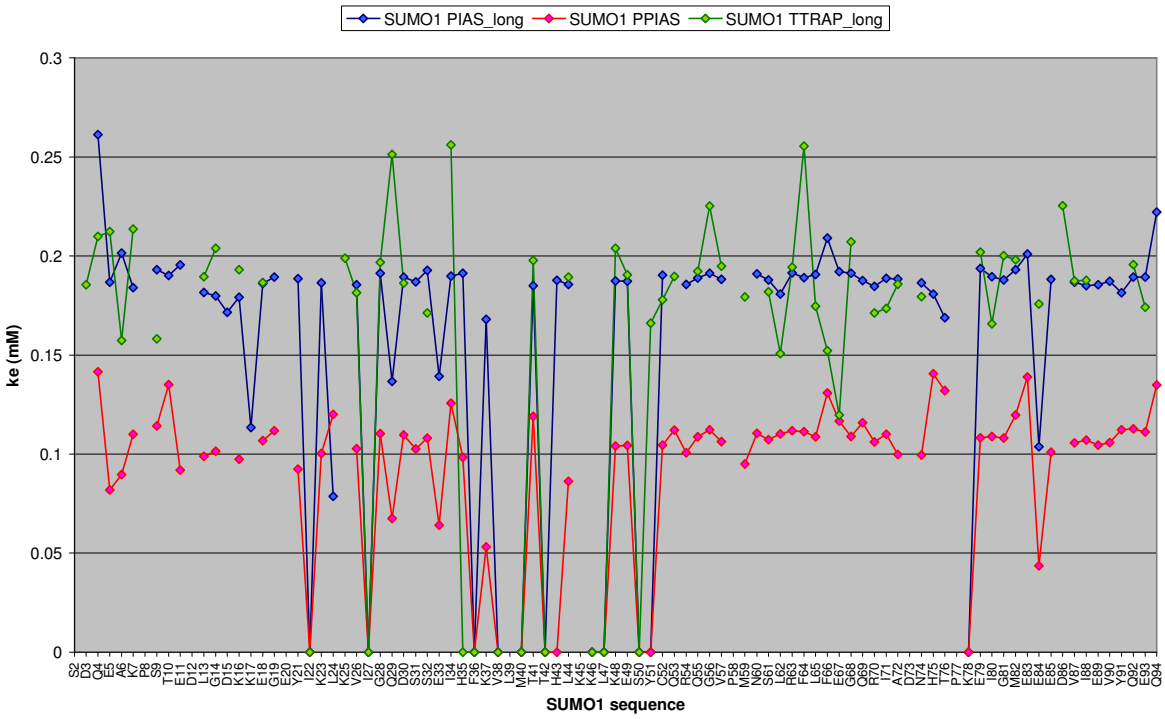
Table 3: K_e (in mM) and Δ_{max} (normalized ppm) values for the amino acids of SUMO1 in different titrations

	SUMO1 PIAS_long		SUMO1 PPIAS		SUMO1 PIAS_short		SUMO1 TTRAP_long	
Residue	ke	delta max	ke	delta max	ke	delta max	ke class	delta max
S2							3	0.003004
D3					0.1854	0.0034	3	0.004105
Q4	0.2612	0.0266	0.1415	0.0528	0.2098	0.187	3	0.187031
E5	0.1867	0.0134	0.0819	0.0114	0.2123	0.002	3	0.001848
A6	0.2015	0.0033	0.0898	0.0028	0.1574	0.0021	3	0.002377
K7	0.184	0.0147	0.11	0.0159	0.2136	0.0056	3	0.005176
P8							3	
S9	0.193	0.0094	0.1142	0.0104	0.1582	0.003		0.002659
T10	0.1901	0.024	0.1351	0.0331			2	0.008835
E11	0.1954	0.0523	0.092	0.0426				
D12							2	0.00231
L13	0.1816	0.0571	0.0989	0.0631	0.1895	0.0108	2	0.011654
G14	0.1798	0.0365	0.1015	0.0424	0.2039	0.0123	2	0.012048
D15	0.1715	0.0518					3	0.002941
K16	0.1791	0.0507	0.0974	0.0572	0.1931	0.0099	2	0.009294
K17	0.1134	0.006					3	0.002517
E18	0.1861	0.0238	0.1068	0.0403	0.1866	0.0032	2	0.003672
G19	0.1893	0.0606	0.1119	0.0674			3	0.004867
E20							2	
Y21	0.1886	0.034	0.0924	0.0447				
I22	0.000003		0.000003		0.000003	0.0374	3	0.034084
K23	0.1865	0.1019	0.1004	0.1433				
L24	0.0788	0.0395	0.12	0.0277			3	
K25					0.1988	0.0056	3	0.005552
V26	0.1855	0.0676	0.1028	0.0664	0.1814	0.0042	3	0.002941
I27	0.000003		0.000003		0.000003	0.0103	2	0.008599
G28	0.1913	0.0374	0.1104	0.0301	0.1968	0.0074	1	0.007917
Q29	0.1367	0.0157	0.0675	0.0144	0.2513	0.009	2	0.007605
D30	0.1893	0.0512	0.1098	0.0559	0.1862	0.0054	3	0.004698
S31	0.1869	0.0546	0.1027	0.0635			2	0.008094
S32	0.1927	0.032	0.1081	0.0344	0.1712	0.0031	3	0.003532
E33	0.1393	0.0369	0.0641	0.0382				
I34	0.1898	0.022	0.1258	0.029	0.2561	0.0026	2	0.002143
H35	0.1912	0.0917	0.0984	0.0948	0.000003		0	0.007978
F36	0.000003		0.000003		0.000003	0.0302	1	0.034574
K37	0.1681	0.0852	0.0532	0.0978				
V38	0.000003		0.000003		0.000003			
L39								
M40	0.000003		0.000003		0.000003	0.0154	2	0.011994
T41	0.185	0.044	0.1191	0.047	0.1978	0.0035	1	0.003056
T42	0.000003		0.000003		0.000003		2	
H43	0.1877	0.1262	0.000003				3	
L44	0.1856	0.0675	0.0863	0.0681	0.1893	0.01	2	0.007765
K45								
K46	0.000003		0.000003		0.000003		3	
L47	0.000003		0.000003		0.000003	0.0208	1	
K48	0.1874	0.0927	0.1041	0.102	0.2039	0.0099	2	0.008553
E49	0.1873	0.101	0.1044	0.1051	0.1904	0.0116	1	0.01147

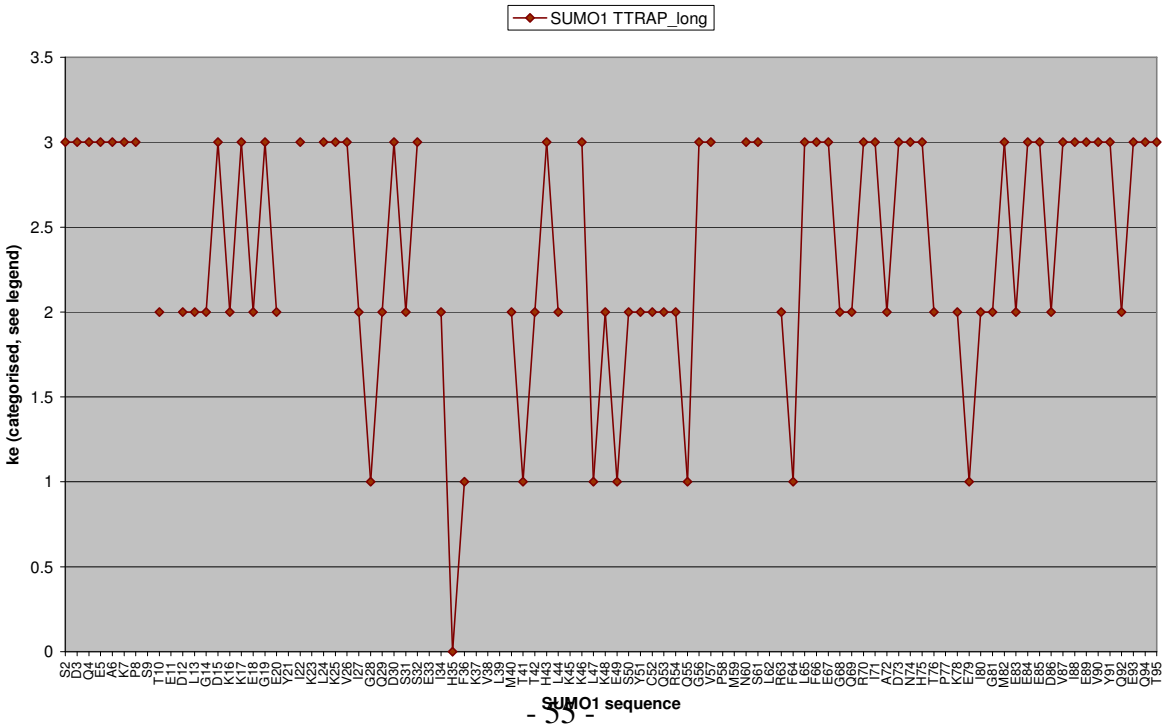
S50	0.000003		0.000003		0.000003	0.0276	2	0.026714
Y51	0.000003		0.000003		0.1661	0.0096	2	0.01517
C52	0.1903	0.0999	0.1046	0.1114	0.1778	0.0059	2	0.007509
Q53			0.1121	0.0579	0.1896	0.0044	2	0.004073
R54	0.1856	0.0613	0.1007	0.0736			2	0.004158
Q55	0.1889	0.051	0.1087	0.0588	0.1923	0.0078	1	0.007765
G56	0.1913	0.0257	0.1123	0.0262	0.2253	0.005	3	0.003978
V57	0.1882	0.0602	0.1063	0.0637	0.1948	0.0061	3	0.005508
P58								
M59			0.095	0.1588	0.1794	0.0153		0.015949
N60	0.191	0.0294	0.1106	0.0386			3	0.007916
S61	0.1878	0.025	0.1073	0.0325	0.1819	0.0034	3	0.004027
L62	0.1808	0.0166	0.1103	0.0124	0.1507	0.005		0.005
R63	0.1914	0.0255	0.1119	0.0224	0.1944	0.0066	2	0.006008
F64	0.189	0.0371	0.1114	0.0422	0.2554	0.0089	1	0.007002
L65	0.1907	0.0446	0.1087	0.0542	0.1746	0.0039	3	0.004473
F66	0.2091	0.0074	0.1309	0.0112	0.1522	0.0023	3	0.002024
E67	0.1921	0.0173	0.1167	0.0185	0.1198	0.0021	3	0.001386
G68	0.1913	0.0391	0.1089	0.0485	0.2071	0.0055	2	0.005386
Q69	0.1876	0.0144	0.1159	0.0375			2	
R70	0.1846	0.0403	0.1062	0.0439	0.1713	0.0045	3	0.004085
I71	0.1887	0.0516	0.1101	0.0652	0.1735	0.0051	3	0.00604
A72	0.1884	0.0674	0.0999	0.0761	0.1857	0.007	2	0.006071
D73							3	
N74	0.1865	0.1088	0.0995	0.145	0.1795	0.0151	3	0.014777
H75	0.1808	0.0125	0.1406	0.025			3	0.004411
T76	0.1689	0.1708	0.1321	0.191			2	0.007136
P77								
K78	0.000003		0.000003				2	0.001967
E79	0.1937	0.0303	0.1083	0.0346	0.202	0.0028	1	0.002377
I80	0.1895	0.0445	0.1089	0.0493	0.1657	0.0061	2	0.006836
G81	0.1878	0.0489	0.1082	0.06	0.2002	0.0067	2	0.006477
M82	0.1931	0.0339	0.1197	0.0434	0.198	0.0058	3	0.00515
E83	0.2009	0.0579	0.139	0.079			2	0.018712
E84	0.1037	0.0057	0.0436	0.0058	0.1757	0.0014	3	0.001386
E85	0.1882	0.1133	0.1011	0.1319			3	0.007529
D86					0.2255	0.0047	2	0.004411
V87	0.1867	0.0738	0.1057	0.0789	0.1874	0.0071	3	0.005937
I88	0.1849	0.0849	0.1072	0.0889	0.1877	0.0077	3	0.008381
E89	0.1854	0.02	0.1045	0.0243			3	0.004085
V90	0.1873	0.054	0.1058	0.0614			3	0.003035
Y91	0.1815	0.0362	0.1124	0.04			3	0.004127
Q92	0.1894	0.0335	0.1128	0.0343	0.1957	0.0077	2	0.007352
E93	0.1894	0.0341	0.1112	0.0353	0.1742	0.0053	3	0.005343
Q94	0.2221	0.0146	0.1349	0.0167			3	0.002203
T95			0.0418	0.0065			3	0.009733

Fig. 31: a) Values of k_e along the sequence of SUMO1 in titrations of SUMO1 with PIAS derived peptides b) The binding of TTRAP_long to SUMO1 being to weak to calculate k_e values with an acceptable accuracy, the titration curves have been divided in 4 classes. 0 corresponds to line broadening (low k_e), 1 to titration curves that appear to reach saturation, 2 to titration curve that do not reach saturation and 3 to amino acid that are not affected by TTRAP_long binding. c) values of Δmax along the sequence of SUMO1 in the different titrations. In the case of the titration with TTRAP, the value of Δ at the highest peptide concentrations are given instead of Δmax, since very few amino acids seem to come close to saturation.

a)



b)



c)

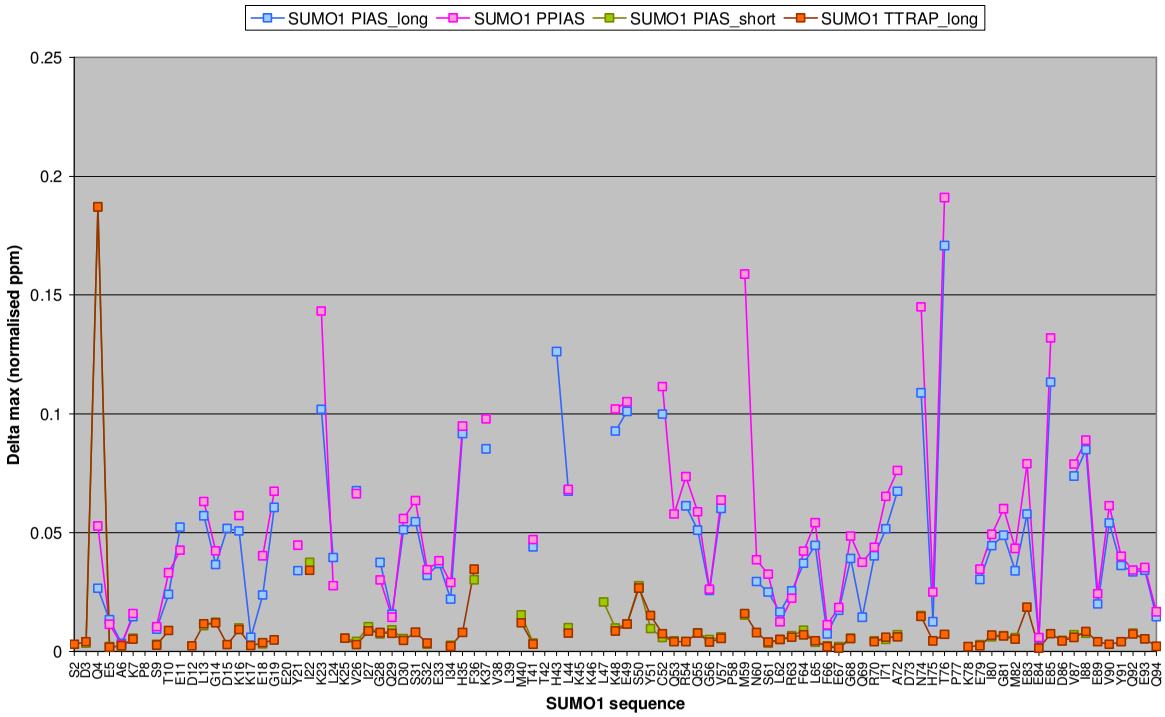

SUMO1 PIAS_long
SUMO1 PPIAS
SUMO1 PIAS_short
SUMO1 TTRAP_long
Delta max (normalised ppm)
0.25
0.2
0.15
0.1
0.05
0
SUMO1 sequence

fig. 32: Values of k_e represented on the structure of SUMO1 (PDB code 1A5R, [Bayer *et al*, 1998]). Color coding is explained in text. High values of ke are in blue, average ones in pink, lower ones in orange. Amino acids with very low values of k_e are in lime green. All views are from the same angle. Ribbon representations are on the left, surface representation of the same on the right. From top to bottom, the values represented are those obtained with PIAS_long, PPIAS, PIAS_short and TTRAP_long.

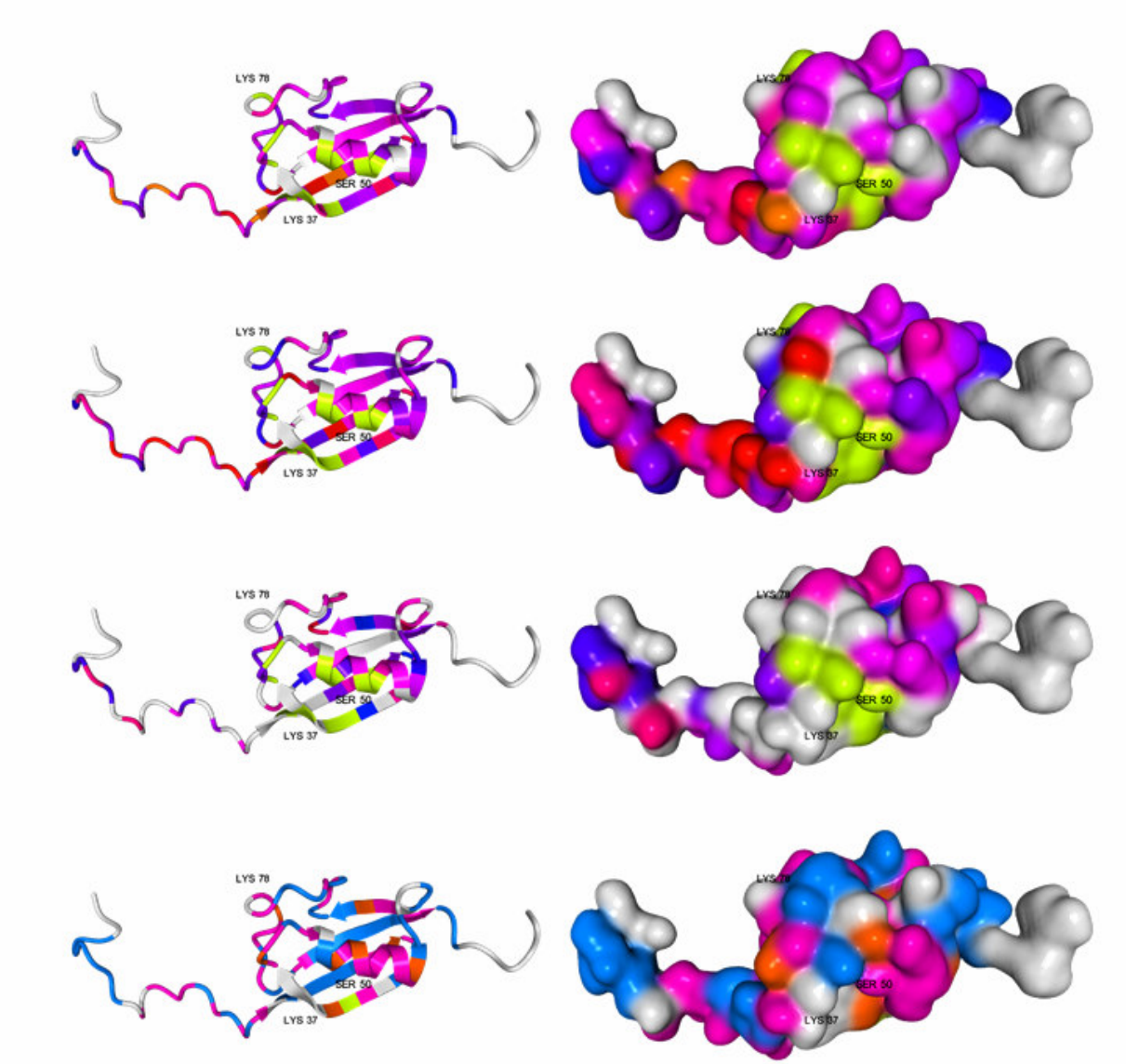


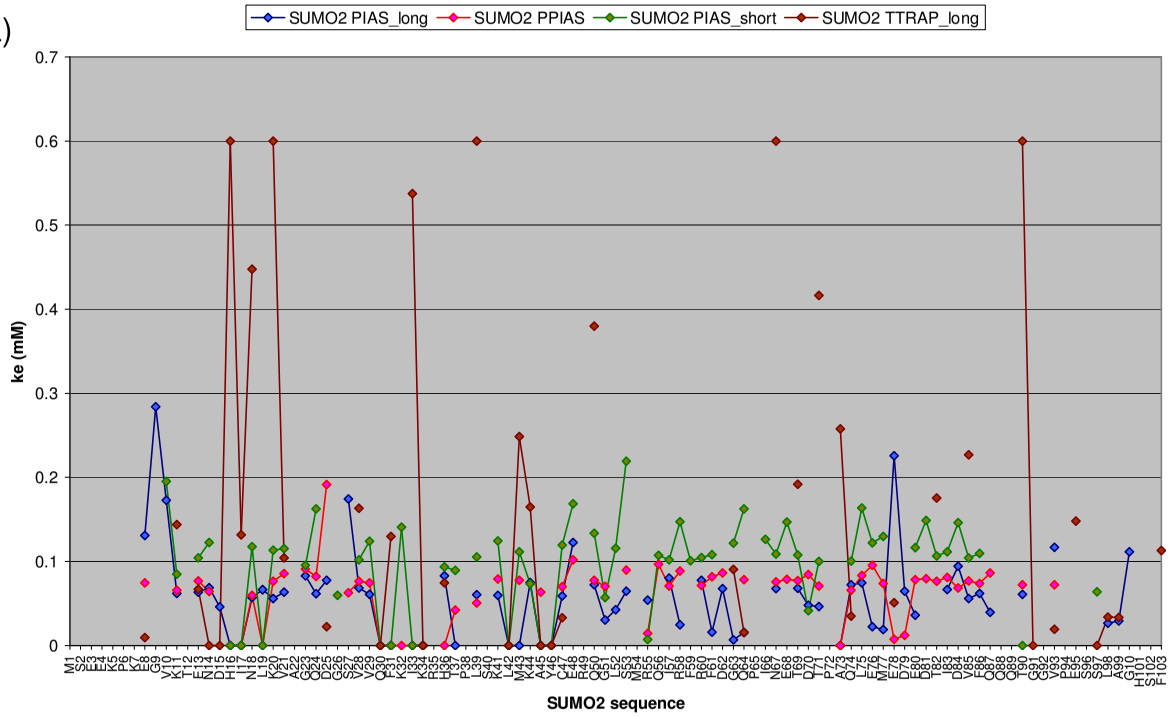
Table 4: K_e (mM) and Δ_{max} (normalized ppm) values for the amino acids of SUMO2 in different titrations

	SUMO2 PIAS_long		SUMO2 PPIAS		SUMO2 PIAS_short		SUMO2 TTRAP_long	
Residue	ke	delta max	ke	delta max	ke	delta max	ke	delta max
E8	0.1306	0.0092	0.0745	0.0047			0.0092	0.0053
G9	0.2837	0.0069						
V10	0.1728	0.0052			0.195	0.0042		
K11	0.062	0.0219	0.0655	0.0137	0.0847	0.0127	0.1436	0.009
T12								
E13	0.0635	0.0262	0.0768	0.0354	0.1042	0.0284	0.0668	0.0126
N14	0.0687	0.1103	0.0641	0.0993	0.1223	0.1127	0.000002	0.0223
D15	0.0459	0.0376					0.000002	0.0452
H16	0.000002		0.000002		0.000002		2.6356	0.1301
I17	0.000002	0.2942	0.000002		0.000002		0.1317	0.0352
N18	0.0568	0.1843	0.0594	0.193	0.1174	0.1595	0.4477	0.0894
L19	0.0662	0.2486	0.000002		0.000002	0.1552		
K20	0.0558	0.1073	0.0764	0.109	0.1134	0.1485	0.935	0.0262
V21	0.0635	0.0591	0.0855	0.0655	0.1152	0.0469	0.1042	0.0095
A22								
G23	0.0828	0.0147	0.0907	0.0225	0.0952	0.0188		
Q24	0.0615	0.0285	0.0821	0.0267	0.1624	0.0229		
D25	0.0777	0.0123	0.1914	0.0076			0.0221	0.0068
G26					0.0594	0.0035		
S27	0.1744	0.0111	0.0626	0.0045				
V28	0.0683	0.027	0.0762	0.0304	0.1018	0.0396	0.1631	0.0068
V29	0.0608	0.0618	0.0745	0.067	0.1239	0.0641		
Q30	0.000002		0.000002		0.000002		0.000002	
F31	0.000002	0.2115	0.000002		0.000002		0.1296	0.0152
K32			0.000002		0.1408	0.1175		
I33	0.000002	0.2762	0.000002		0.000002		0.5376	0.1324
K34	0.000002	0.2214	0.000002		0.000002	0.1809	0.000002	
R35								
H36	0.0827	0.1269	0.000002		0.0937	0.0954	0.0744	0.092
T37	0.000002	0.2314	0.0419	0.2293	0.0891	0.1743		
P38								
L39	0.0604	0.1796	0.0505	0.1746	0.1051	0.1552	4.4645	0.1895
S40								
K41	0.0596	0.1023	0.0789	0.1296	0.1242	0.049		
L42	0.000002				0.000002		0.000002	
M43	0.000002	0.1328	0.0774	0.1311	0.1115	0.1155	0.2487	0.0412
K44	0.0749	0.0649			0.0733	0.0761	0.1646	0.0255
A45	0.000002	0.1963	0.0634	0.1781	0.000002	0.1834	0.000002	0.0685
Y46	0.000002				0.000002		0.000002	0.0048
C47	0.0588	0.0719	0.0698	0.0783	0.1195	0.0772	0.0329	0.0218
E48	0.1224	0.0182	0.102	0.0235	0.1687	0.0391		
R49								
Q50	0.0721	0.0373	0.0773	0.0374	0.1336	0.0336	0.3798	0.0278
G51	0.0301	0.0246	0.0702	0.014	0.0568	0.0226		
L52	0.0425	0.0065			0.1155	0.0123		
S53	0.0643	0.007	0.0897	0.0057	0.219	0.0396		
M54								

R55	0.054	0.0345	0.0142	0.0333	0.007	0.0144		
Q56			0.0965	0.0072	0.1072	0.0062		
I57	0.0802	0.013	0.0706	0.0148	0.1018	0.0131		
R58	0.0245	0.0141	0.0885	0.011	0.1472	0.013		
F59					0.1006	0.0173		
R60	0.0774	0.0362	0.0712	0.0435	0.1048	0.0521		
F61	0.0158	0.0369	0.0817	0.0288	0.108	0.0311		
D62	0.0678	0.0993	0.0864	0.0093				
G63	0.0068	0.0423			0.1216	0.0077	0.0903	0.0305
Q64	0.0154	0.0225	0.0784	0.0095	0.1623	0.0077	0.0153	0.0194
P65								
I66					0.1262	0.0675		
N67	0.0677	0.0803	0.0758	0.075	0.1087	0.0794	3.6735	0.1734
E68			0.0786	0.0716	0.1467	0.0125		
T69	0.0679	0.1065	0.0772	0.1052	0.1076	0.0667	0.1918	0.0201
D70	0.0479	0.0193	0.0842	0.0143	0.0414	0.027		
T71	0.0464	0.0973	0.0708	0.0889	0.1	0.0815	0.4165	0.0649
P72								
A73	0.000002		0.000002				0.2575	0.0771
Q74	0.0722	0.0319	0.0658	0.0155	0.1004	0.0149	0.0348	0.0326
L75	0.0745	0.0418	0.0832	0.0439	0.1636	0.0542		
E76	0.0225	0.0419	0.0956	0.0243	0.122	0.0168		
M77	0.0189	0.0085	0.0737	0.01	0.1296	0.0165		
E78	0.2258	0.1402	0.0076	0.0117			0.0509	0.0889
D79/F59	0.0647	0.016	0.0121	0.0211				
E80	0.0361	0.0842	0.0784	0.0712	0.1165	0.0555		
D81			0.0792	0.0185	0.1488	0.0189		
T82			0.0764	0.0115	0.1064	0.0258	0.1752	0.0098
I83	0.0666	0.0824	0.081	0.0775	0.1114	0.0701		
D84	0.094	0.0244	0.0683	0.0271	0.146	0.0232		
V85	0.0558	0.0973	0.0767	0.0883	0.1042	0.0912	0.2268	0.0245
F86	0.0621	0.0764	0.0737	0.0849	0.1096	0.0912		
Q87	0.0394	0.0055	0.0864	0.0102		0.0284		
Q88								
Q89								
T90	0.0607	0.2064	0.0722	0.1957	0.000002	0.1829	1.1078	0.1659
G91							0.000002	
G92								
V93	0.1167	0.0841	0.0722	0.0079			0.0193	0.179
P94								
E95		0.0192					0.1482	0.0104
S96								
S97					0.0638	0.1698	0.000002	
L98	0.0264	0.0139					0.0337	0.0123
A99	0.0289	0.0119					0.0333	0.0119
G100	0.1113	0.0137						
H101								
S102								
F103							0.1127	0.0185

Fig. 33: a) Values of k_e along the sequence of SUMO2 in titrations of SUMO2 with the different peptides In the case of the TTRAP titration, all k_e values higher than 0.6 were set to 0.6 for sake of legibility. b) values of Δmax along the sequence of SUMO2 in the different titrations.

a)



b)

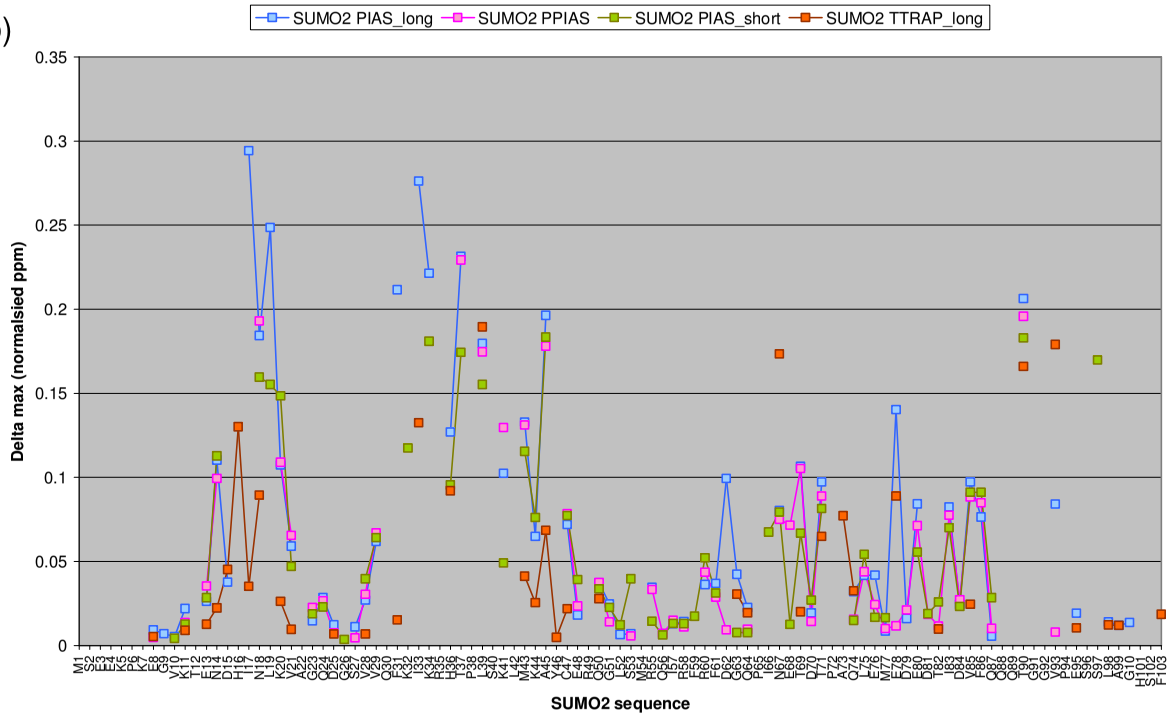
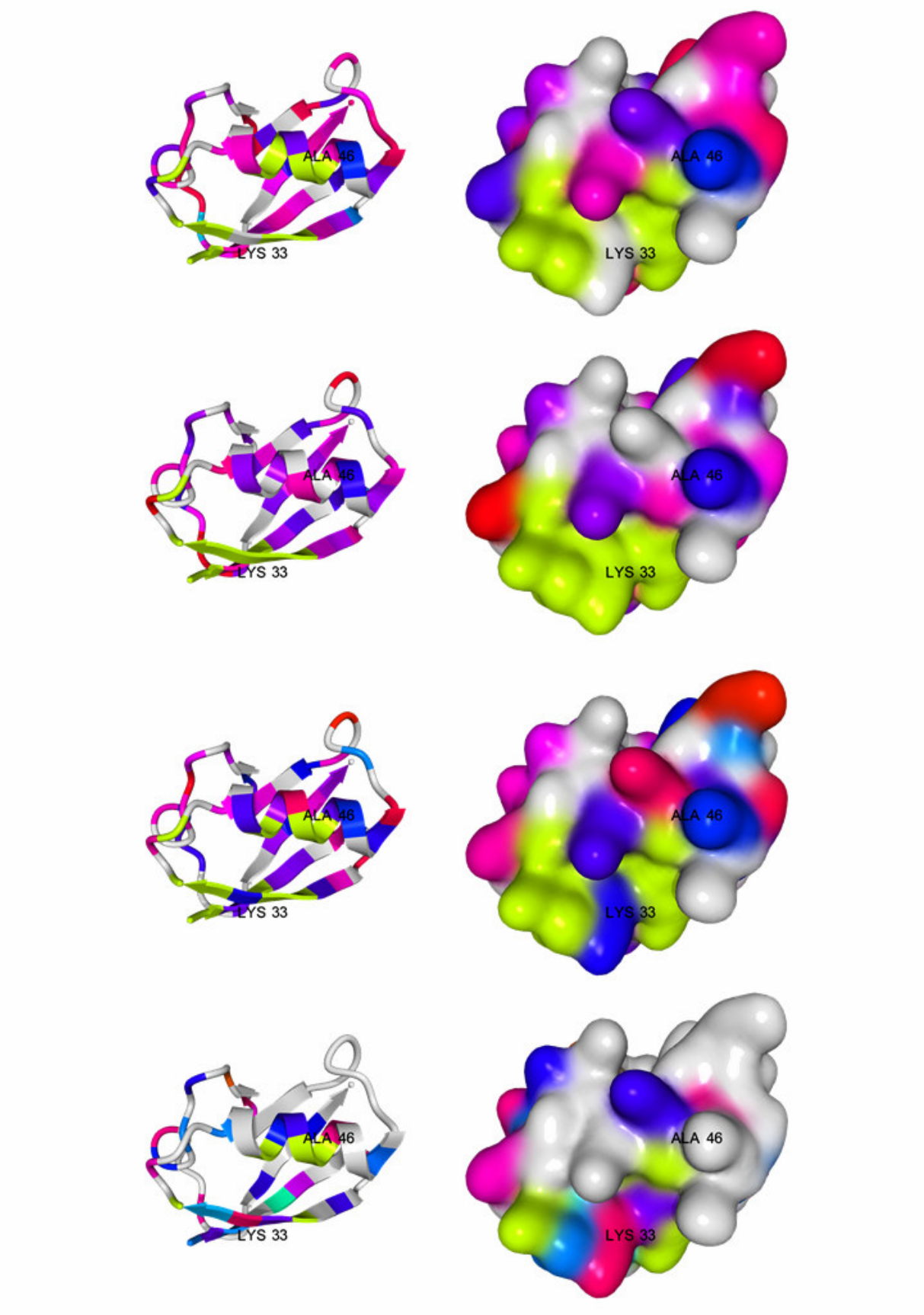


Fig. 34: Values of k_e represented on the structure of SUMO2 (PDB code 1WM3, [Huang *et al*, 2004]). Color coding is explained in text. High values of ke are in blue, average ones in pink, lower ones in orange. Amino acids with very low values of k_e are in lime green. All views are from the same angle. Ribbon representations are on the left, surface representation of the same on the right. From top to bottom, the values represented are those obtained with PIAS_long, PPIAS, PIAS_short and TTRAP_long.



8. Binding to PIAS and TTRAP derived peptides cause similar changes in the environment of the amino acids of SUMO1 and of SUMO2

The comparison of the graphical representation of Δ_{max} along the sequence of SUMO2 for the different titrations shows that SUMO2 responds in the same way to any of the tested peptides. High Δ_{max} and low k_e values are found in the same regions of the protein: at the beginning of the β1 strand between His 16 and Lys 20, between the second half of the β2 stand and the first half of the α-helix from Phe 31 to Ala 45, and between Ile 66 and Phe 86 in the loop region separating the β3- and β4-strands (fig. 35a). An overlay of the HSQC spectra obtained with SUMO2 at the end of the different titrations show that nearly all SUMO2 N and HN resonances are at the same position when SUMO2 is saturated with PIAS_long, PPIAS and PIAS_short. Similar

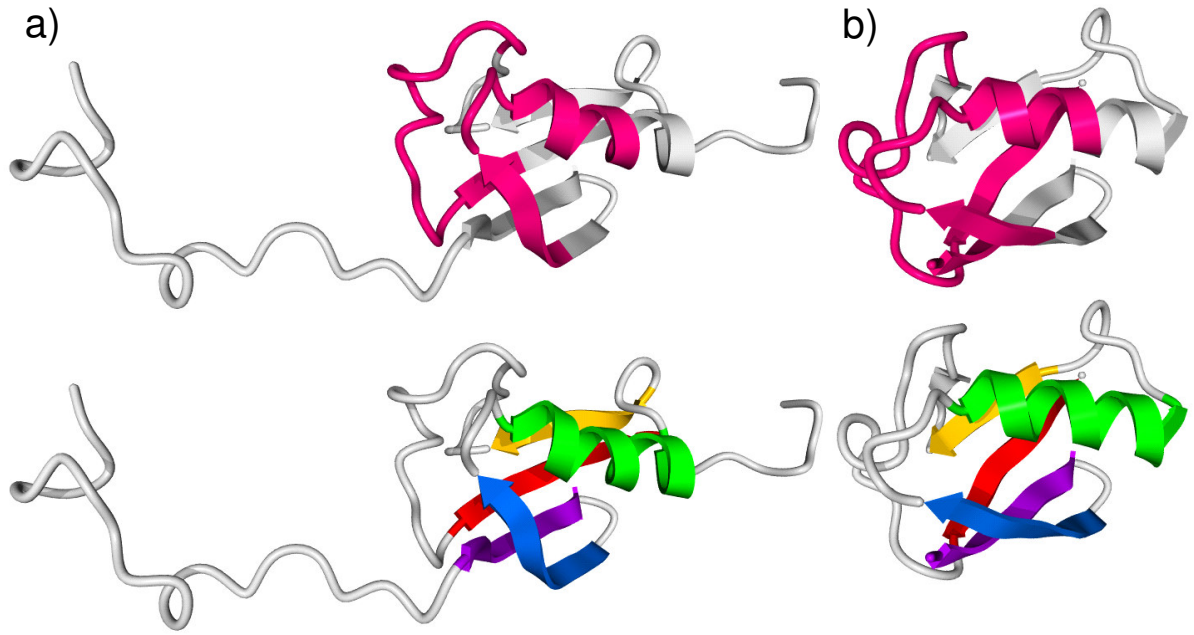


Fig. 35: regions of SUMO1 (a) and SUMO2 (b) with amino acids that have high Δ_{max} and low k_e value are represented in pink on the upper part of the figure. The lower part shows the same structures (1A5R and 1WM3) from the same point of view with the coloring scheme presented in the introduction.

observations can be made with SUMO1: low values of k_e occur between His 35 and Tyr 51, values of Δ_{max} tend to be higher between His 35 and Gln 53 and between Ala 72 and Ile 88 (fig. 35b). HSQC spectra are similar at saturation with PIAS_long and PPIAS, and HSQC spectrum taken at saturation with PIAS_short is similar to that taken at the highest concentration of TTRAP_long. A comparison of the localization

of the amino acids in intermediate exchange (i.e. those that are most affected by the peptide binding) in the different titrations (table 5) shows again a strong similarity between the effect of the binding of the different peptides. In this table, those amino acids are divided in two classes, depending on whether their HSQC peak broadens significantly or completely disappears during the titration. Given that the differences in resonance frequencies between the free and bound protein environments are the same, in the first case (fast intermediate exchange), the exchange is faster than in the second (slow intermediate exchange), corresponding to larger k_e values in the range of approximation. This distinction shows that the different peptides do not bind with the same affinity to SUMO.

Those amino acids are clearly grouped within and around the second β-strand and the α-helix of SUMO. The representation of those amino acids on SUMO1 and 2 structures (fig. 36) show that they concentrate in a patch of SUMO structure, which is the binding site for the SIM on SUMO. However, not all of those amino acids are actually part of the binding site. The surface representations of SUMO1 and 2 in fig.

SUMO1-PIAS_long	SUMO1-PPIAS	SUMO1-PIAS_short	SUMO1-TTRAP_long	SUMO2-PIAS_long	SUMO2-PPIAS	SUMO2-PIAS_short	SUMO2-TTRAP_long
							N14 (LB)
							D15 (LB)
				H16	H16	H16	
I22	I22	I22 (LB)		I17	I17	I17	
					L19 (LB)	L19 (LB)	
I27	I27	I27 (LB)					
H35 (LB)	H35 (LB)	H35 (LB)	H35 (LB)	Q30	Q30	Q30	Q30 (LB)
F36	F36	F36 (LB)		F31	F31	F31	
	K37 (LB)				K32		
V38	V38	V38 (LB)		I33	I33	I33	
				K34	K34	K34 (LB)	K34 (LB)
M40	M40	M40 (LB)			H36		
				T37			
T42	T42	T42					
	H43 (LB)						
K46	K46	K46			K41 (LB)		
L47	L47	L47		L42	L42	L42	L42 (LB)
				M43 (LB)			
S50	S50	S50		A45	A45 (LB)	A45 (LB)	A45 (LB)
Y51	Y51			Y46	Y46	Y46 (LB)	Y46 (LB)
K78	K78 (LB)			A73	A73		
						T90 (LB)	
							G91 (LB)
					S97		S97 (LB)
11 (1)	10 (4)	4 (6)	0 (1)	11 (1)	12 (3)	6 (5)	0 (9)

Table 5: Amino acids in intermediate exchange upon binding of SUMO with PIAS and TTRAP derived peptide. The amino acids for which the line broadening is not important enough to cause an apparent disappearance of the peak (intermediate fast exchange) are signaled by the mention "LB". Amino acids constituting the binding site described in the text are highlighted in Salmon. Positions equivalent in SUMO1 and 2 (see introduction, fig. 2) are in the same row in this table. In the last row, the number of amino acids in intermediate and in intermediate fast (in parenthesis) exchange is indicated.

36 shows that only some of them actually emerge at the surface of SUMO, whereas others are completely buried beneath it and cannot therefore make direct interaction with the peptides.

By proceeding so, using Yasara to generate the molecular surface for SUMO1 and -2 structures, the SIM binding site of SUMO1 is found to consist of His35, Phe36,

Lys37, Val38, Thr42, His43, Lys46 and Ser50. Several amino acids that are in intermediate exchange in one or the other titration of SUMO1 can be excluded from the binding site. Lys78 is far away from any other amino acid in intermediate exchange, and is itself in intermediate exchange only in the PPIAS titration. It is therefore not part of the SIM binding site. Ile22, Met40, Leu47 and Tyr51present

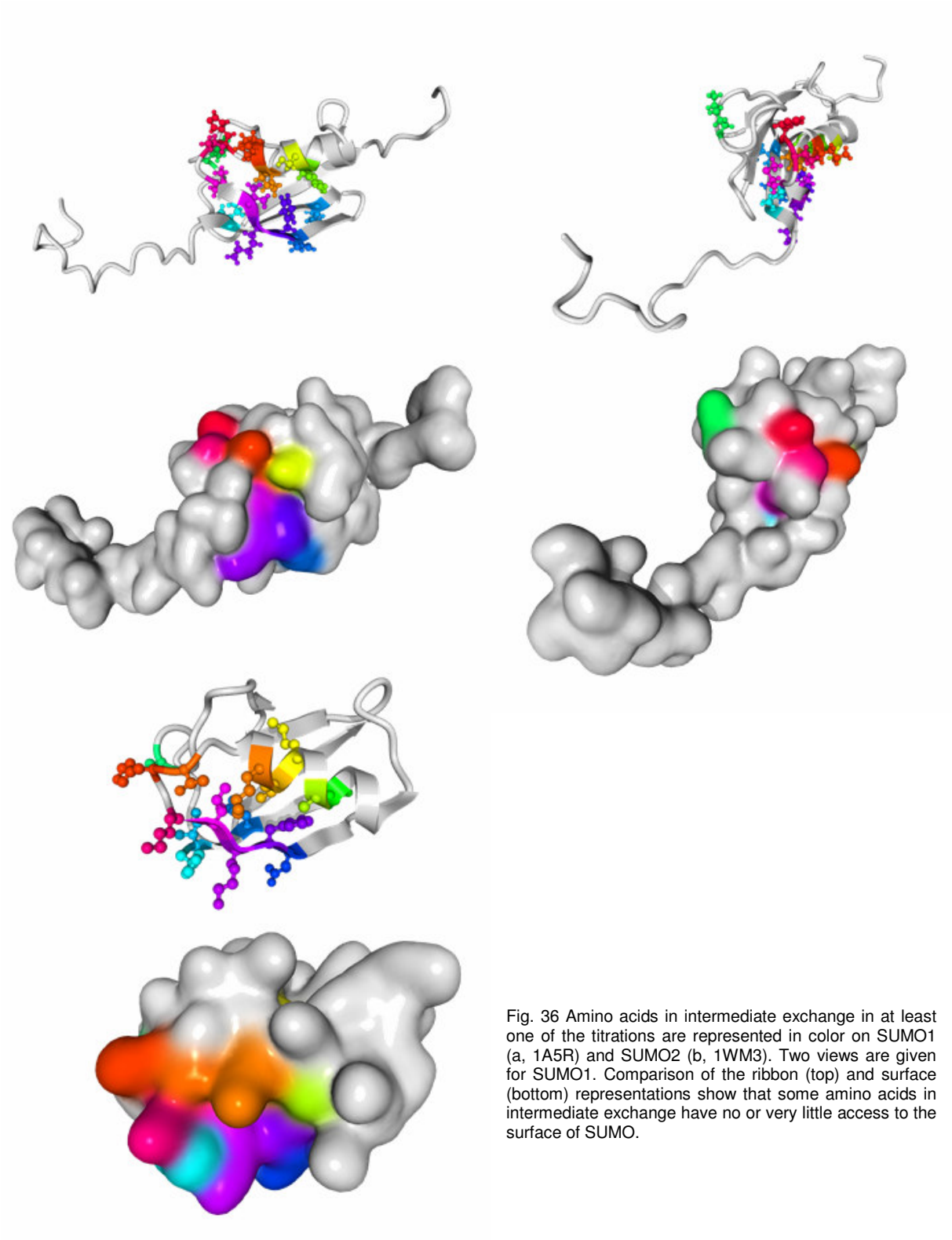


Fig. 36 Amino acids in intermediate exchange in at least one of the titrations are represented in color on SUMO1 (a, 1A5R) and SUMO2 (b, 1WM3). Two views are given for SUMO1. Comparison of the ribbon (top) and surface (bottom) representations show that some amino acids in intermediate exchange have no or very little access to the surface of SUMO.

nothing or very little of themselves at the surface of SUMO1. They are therefore not part of the SIM binding site of SUMO. The case of Ile27 is difficult to decide. It lies out of the patch formed by the SIM binding site amino acids, but still very close to it. The available data does not allow deciding whether it is part of the SIM binding site, or is merely indirectly influenced by the peptide binding. Lys39 is probably part of the SUMO binding site, but its HSQC peak overlaps the Glu33 peak in the free protein state, making it impossible to decide whether Lys39 is in intermediate exchange or not.

The same procedure shows that the SIM binding site of SUMO2 consists of His16, Gln30, Phe31, Lys32, Ile33, Lys34, His36, Lys41 and Ala45. All other amino acids in intermediate exchange have side chains that are part of the hydrophobic core of SUMO2 directly beneath the SIM binding site. This is recapitulated in table 5.

Discussion

1. Studying the properties of peptide binding sites on protein based on NMR titration data

The titrations performed for SUMO1 and SUMO2 give us two pieces of information on the behavior of most amide groups in those proteins. The first is exchange constant k_e, which tells how much longer this amino acid spends in one environment than in the other. It can often be calculated precisely in the cases where the considered amino acids are in fast exchange, and estimated within an order of magnitude when the considered amino acid is in intermediate exchange. The second information is the magnitude of the chemical shift change (Δ_{max}) caused by the peptide binding. There is no method available to quantify the structural change undergone by a protein based on the change of chemical shift of its amino acids. It can be however safely assumed that, in general, a particularly large change of chemical shift for an amino means that this amino acid undergoes particularly great changes in environment.

Both types of information can be combined to determine which amino acids of a protein are involved in binding a peptide. Both types of information are to be dealt with cautiously. A small value of k_e for an amino acid –meaning that it stays preferentially in the "bound protein" environment- or a high Δ_{max} –meaning the "bound protein" and "free protein" environments are very different for this amino acid- do not necessarily mean that this amino acid makes direct contacts with the peptide. It can as well make a strong interaction with other amino acids contacting the peptide, or be under the influence of overall structural rearrangements caused by the peptide binding. Therefore the k_e and Δ_{max} values must be carefully analyzed in the context of the whole protein to reach conclusions on the details of peptide binding. However it can be done better by this method better than by any other existing one because it is the only possibility to observe the behavior of single amino acids in solution.

2. The SIM binding surface of SUMO is a "universal plug" through which proteins with SIM can interact with SUMO

As exposed in the results, PIAS- and TTRAP derived peptides bind to the same surfaces of SUMO1 and -2, and those surfaces are on similar locations in

SUMO1 and -2. It can be concluded from these similarities that PIAS and TTRAP derived peptides bind in the same manner respectively to each other to SUMO1 and 2. It is important to note that the two regions observed to be particularly affected by PIAS and TTRAP derived peptide binding are in the folded part of SUMO, which is absolutely identical between SUMO2 and SUMO3. There should therefore be no substantial differences between the binding of those peptides to SUMO2 and 3. This binding site corresponds well to the one proposed by [Song *et al*, 2004] for the PIAS_long peptide.

The PIAS and TTRAP derived peptides used in this study being representative of the SUMO Interacting Motif (SIM) found in many proteins, it can be safely assumed that all those proteins bind to all three SUMO isoforms in a similar way, differences in the details of the binding accounting for the different isoform specificities and affinities of those proteins, as will be discussed in the following sections. Thus, the interaction binding surface of SUMO for PIAS and TTRAP appears as a "universal plug" through which SUMO can interact with any protein that contains a SIM. The observation by [Chupreta *et al*, 2005] of a surface determining the inhibitory role of SUMO in DNA transcription that corresponds to the SIM binding site described here confirms the universality of the SIM/SIM binding site interface for mediating SUMO interactions.

The hydrophobic core of the SIM is the primary determinant of the binding to SUMO. Without it, the binding to SUMO is completely eliminated, as shown by the mutation experiments made by Christina Hecker. It can therefore be concluded that the hydrophobic core contacts the SIM binding site of SUMO. Since it has a sequence typical of a β-strand and that the SIM binding site on SUMO is situated at the side of SUMO's β-sheet, it is probable that the hydrophobic core of the SIM forms an intermolecular β-sheet with the β2-strand of SUMO, somewhere between Glu33 and Lys39. Two structures published after this work was done (the structures of the SUMO1/RanGap1/RanBP2/Ubc9 complex [Reverter & Lima, 2005] and of the sumoylated Thymine DNA Glycosylase (TDG) [Baba *et al*, 2005]) comfort this hypothesis. In both structures, SUMO1 is in complex with another protein (RanBP2 and TDG respectively) through an intermolecular β-sheet involving SUMO1's β2 strand and a β-strand situated in an unstructured region of the other protein. In the case of RanBP2, this β-strand contains the SIM's hydrophobic placed aside the part

of SUMO1's β2-strand between Val38 to His35 giving a case of a SIM binding to SUMO according to the predictions based on the present analysis of NMR data.

3. Mechanisms by which different affinities are observed within a protein

Analysis of participation to SUMO's surface of the amino acids in intermediate exchange in the different titrations has shown that some amino acids that are strongly affected by peptide binding can not make direct contacts with the peptide itself (Ile22, Met40, Leu47 and Tyr51in SUMO1; Leu 42, Met 43 and Tyr 46 in SUMO2). The strong effect of peptide binding on them is indirect: those amino acids have sidechains buried in the hydrophobic core of SUMO, in immediate proximity of the SIM binding site. The structural changes in the SIM binding site caused by peptide binding are changes of environment for the amino acids situated directly beneath it. The mechanism responsible for this also explains the surprising fact that different k_e values are observed within a protein when it binds to a ligand. *A priori*, if the ligand binds at a certain frequency, and the protein-ligand complex has a certain lifetime, then each of the amino acids of the protein should change environment with this frequency and lifetimes. One has, however, to take into account that a protein is not a rigid object. Each of the amino acids of a protein is able to move relatively to the others, and to adopt a conformation that is stable in its environment. Therefore, a change of position of an amino acid X in response to ligand binding does not necessarily modify the position of all others, but merely constitutes a change of environments for the neighboring amino acids. An amino acid Y in the neighborhood of X has now two possible conformations: the "normal" conformation and the conformation adapted to the modified environment. Y will exchange between both with frequency and lifetimes depending on the stability of both conformations *and* on the rate at which X's conformation is modified, *not* on that rate only. This accounts for the fact that the k_e varies from amino acid to amino acid in the titrations of SUMO with PIAS and TTRAP derived peptides. This also explains why some amino acids are in intermediate exchange upon titration with PIAS- and TTRAP derived peptides although they can make no direct contacts with those peptides: those amino acids have side chains buried in the hydrophobic core of SUMO directly beneath the SIM binding site, a situation which makes them particularly sensitive to the structural rearangments of the SIM binding site induced by peptide binding.

4. The PIAS and TTRAP derived peptides bind with different affinities to SUMO1 and SUMO2/3

Until this point, the analysis of the binding of the studied peptides to SUMO has relied on values of the exchange constant k_e for individual amino acids, which indicates the preference that those amino acids have for the "bound protein" environment over the "free protein" environment when the peptide is present. To interpret those results in a manner that is relevant to cellular contexts, one has to quantify the binding of proteins to SUMO in a molecule-wide manner. It is, fortunately, quite easy to have an idea of the K_D (dissociation constant) of a peptide for SUMO when the k_e values are known: SUMO as a whole cannot exchange faster between its bound and free state than its amino acid that does so the slowest. Hence, the value of K_D for a peptide and SUMO is lower than the lowest value of k_e found in SUMO with this peptide. It might, of course be lower: conceivably, the different amino acids in the binding site might interact independently with the peptide. In that case, the situation would be that of an array of binding sites. In such situations, the affinity of the array for the ligand is higher than the affinity of each of the array's elements. There is no practical way to investigate those possibilities using NMR measurements, which limits us to an upper-limit estimation of the K_D of the used peptides for SUMO. However, the fact that the K_D measured by isothermal calorimetry (ITC) by [Song *et al*, 2004] (~6 μM) for SUMO1 with the PIAS_long and PIAS_short peptides are in the same order of magnitude than the k_e measured by NMR for the same peptide, showing that at least in the case of SUMO-SIM binding the k_e values allow a fairly good estimation of the K_D.

The table 5 shows that even though the peptides studied here all bind through the same interface to SUMO they affect SUMO to different extent. The number of amino acids in fast intermediate exchange and in slow intermediate exchange is different in each titration. This means that, even though the estimated k_e values for the binding site amino acids are very similar (varying between 2 and 10 μM in each titration where it could be estimated), there are some differences in the affinity of the different peptides to SUMO.

SUMO1 binds very poorly to TTRAP. No detectable change in SUMO1 HSQC spectra can be seen upon addition of either TTRAP_short_Cside or

TTRAP_short_Nside. Some changes are observed upon addition of TTRAP_long, but are too little to allow a precise estimation of the k_e values. In this titration, only His 35 is observed to be in intermediate fast exchange, meaning that its k_e should be no smaller than 3µM (value observed in the titrations of SUMO1 with PIAS derived peptides). Furthermore the k_e derived from the few TTRAP_long/SUMO titration curves that appear to approach saturation are never smaller than 0,12 mM (values obtained for Gly28 and Glu 49). Hence, the K_D of SUMO1 for TTRAP long must lie somewhere between 3 and 120 µM). Much more extensive changes were observed with all three PIAS derived peptides, and all three cause several SUMO1 amino acids to be in intermediate exchange, meaning that they have higher affinity to SUMO1 than the TTRAP derived peptides. Comparing the numbers of SUMO1 amino acids in slow- and fast intermediate exchange (see table 5) in the titrations of SUMO1 with PIAS derived peptides shows that PPIAS has a higher affinity to SUMO1 than PIAS_long, which in turn binds better to SUMO1 than PIAS_short does. In each case, the lowest measured k_e value is 3µM, corresponding to a K_D in this range. Comparing the k_e values for the amino acids in fast exchange in the different titrations lead to the same classification: in term of affinity to SUMO1, (TTRAP_short_Nside / TTRAP_short_Cside) < TTRAP_long << PIAS_short < PIAS_long < PPIAS.

Comparison of the number of amino acids in slow intermediate exchange and in fast intermediate exchange in the different SUMO2 titrations (see table 5) shows that the PIAS derived peptides have a higher affinity to SUMO2 than the TTRAP derived peptides. TTRAP_long has a higher affinity to SUMO2 than TTRAP_short_Nside and TTRAP_short_Cside, which both bind poorly to SUMO2. There appears to be little difference in the affinity of the three studied PIAS derived peptides to SUMO2, the Kd being in each case around 2µM. Comparing the k_e values for the amino acids in fast exchange in the different titrations lead to the same classification: in term of affinity to SUMO2, (TTRAP_short_Nside / TTRAP_short_Cside) < TTRAP_long < (PIAS derived peptides).

Comparing the binding of each peptide to SUMO1 and -2 is only partly possible. TTRAP_long clearly binds better to SUMO2 than to SUMO1 (several amino acids are in intermediate exchange in the former and only one in the latter upon binding to TTRAP_long), confirming the results obtained by Christina Hecker with

GST-pulldown experiments. Such comparison for the other peptides is not possible, the number of amino acids in intermediate exchange being similar in both isoforms.

5. PIAS_short and PIAS_long bind to SUMO1 by a distinctive 2-steps mechanism

For a (Protein + Ligand → Protein•Ligand) binding, it is expected that the titration curve has the shape presented in fig. 30 for SUMO2 This expected shape is actually observed for the amino acids in fast exchange in all titrations but those of SUMO1 with PIAS_long and PIAS_short. For those two titrations, titration curves have a clearly two stepped shape, as shown in fig. 30. Several reasons can be invoked for this unexpected behavior.

The first possibility is that there are several binding sites for PIAS_long and PIAS_short on SUMO.

The fact that the titration curves obtained for SUMO1 with PPIAS have the expected shape point at two alternative explanations. When SUMO1 titrations with PIAS and PPIAS are compared, it appears that the only significant difference in ke between both is for Lys37, which is in fast exchange with PIAS_long and in fast-intermediate exchange with PPIAS. This shows that the phosphate group of PPIAS likely binds to SUMO1 in the neighborhood of this lysine (the same difference is observed for SUMO2's Lys34 which is equivalent to SUMO1's Lys37, confirming this hypothesis). Examination of SUMO1 and SUMO2 structures in this region show that the end of the β2 strand is much more bent in SUMO1 than is SUMO2, and that SUMO1's Lys39 sidechain points into the binding site (it is coplanar with the β-sheet) whereas the equivalent Lys34 in SUMO2 points away from the SIM binding site (it is perpendicular to the β-sheet). Therefore it can be proposed that the negatively charged phosphate group of PPIAS interacts with the positively charged Lys39 of SUMO1 causing a conformation change favorable for PPIAS binding. The absence of phosphate group in PIAS_long would make this transition more difficult to achieve. A high PIAS_long concentration would be required to maintain SUMO1 in the binding-favorable conformation, producing the observed two-steps titration curve. The conformation Lys34 in SUMO2 being already favorable to PIAS derived peptides binding, whether phosphorylation is present or not on the peptide does not matter

much in their binding. Unfortunately, the observation of the behavior of Lys39, which is critical, is made impossible by HSQC peak overlapping.

The last possibility is that the two-steps titration curves observed for titrations of SUMO1 with PIAS_long and PIAS_short is that those peptides have to compete with the N-terminal tail of SUMO1 for binding to the SIM binding site. The N-terminal tail of SUMO1 is observed to form a βstrand parallel to β2 in the complex between SUMO1 and the E1 enzyme Sae1/Sae2 [Lois & Lima, 2005; this complex is presented in ore detail in the introduction], thus blocking access to the SIM binding site. It can be hypothetized that PPIAS would compete more efficiently with the N-terminal tail of SUMO for the SIM binding site because its phosphate group would allow it to interact better with the lysins (particularly Lys37) of the binding site, making the effect of the competitive binding on the titration curves less noticeable. This hypothesis will have to be tested by investing the effect of removing the N-terminal tail of SUMO1 on PIAS derived peptides binding.

Of the three possible explanations for the two-step binding of PIAS_long and PIAS_short to SUMO1 (multiple binding sites, favorable conformation changes more efficiently induced by PPIAS than by PIAS_long, competition of the peptide with the N-terminal tail of SUMO1), the two latest –which are not exclusive- are the most likely to be relevant. The other option raises two difficulties. There is no evidence of a second binding site for PIAS_long and PIAS_short in the titrations of SUMO1 with those peptides, and it is not likely that PIAS_short has two binding sites for SUMO (it is too small for this). Besides, two binding sites would be likely to increase the affinity of unphosphorylated PIAS derived peptides for SUMO, instead of decreasing it.

6. Contribution of the amino acids surrounding the SIM to SUMO binding

The region of SUMO that binds to the SIM has now been defined. But how does the SIM bind to SUMO, and which roles play its different elements? The apparently obvious way to answer those questions is the production of a structure of a SUMO-SIM complex. This is, however, rather impractical. A crystal structure would, even if crystals could be produced at all, be difficult to interpret because the crystalline arrangement might force the peptide to adopt a particular conformation that is not relevant of its behavior in solution. A NMR structure would require large amounts of isotope labeled synthetic peptide which cost would be prohibitive. This

Conclusion

limits us to the information that can be gained from the effects of the peptide binding on SUMO.

As discussed earlier, the sequence of the SIM and the localization of the binding surface on SUMO show that SIMs bind to SUMO by forming an intermolecular β-strand with the β2 strand of SUMO. This gives two possible relative orientations for the SIM and SUMO. In the case of the binding of PIAS- and TTRAP derived peptides to SUMO, several facts hint at a parallel orientation. First, the binding of those peptide has no particular effect on the C-terminal side of the β2 strand of SUMO, whereas is affects several amino acids in the loop region between the β2 strand and the α-helix. As seen before, SUMO1's Lys37 and the corresponding SUMO2's Lys34 are in intermediate exchange with PPIAs and not with the other peptides, indicating that the phosphate group of PPIAS is in proximity of these lysines when PPIAS binds to SUMO. Finally, the amino acids in the loop between the β3 and β4 strands of SUMO are affected by the binding of PIAS_long and PPIAS, but not by PIAS_short, indicating that the C-terminal part of PIAS_long and PPIAS –which is absent in PIAS_short- interacts with this part of SUMO. Those three observations support the conclusion that the core of PIAS-derived peptides is parallel to the β2 strand of SUMO in the PIAS-SUMO complex. It is more delicate to draw conclusions for TTRAP, but the fact that several amino acids in the loop between the β3 and β4 strands of SUMO2 have relatively high affinity for TTRAP_long points to a parallel orientation of this peptide and the β2 strand of SUMO.

The large decrease in affinity of PIAS and TTRAP derived peptides when the amino acids surrounding the SIM are absent shows that those amino acids should play a role in the SIM binding to SUMO. The binding of the PIAS_long and PPIAS, but not of PIAS_short, affects the loop region between SUMO's β3 and β4 strands (particularly Lys78 of SUMO1 and Ala73 of SUMO2, which are in intermediate exchange with those peptides). It can therefore be concluded that the amino acids in C-term of the hydrophobic core of SIM's PIAS contributes to SUMO binding by interacting with this loop.

SUMO proteins appear to be in most cases tags used to mediate binding between their target protein and an interaction partner. This mechanism and the diversity of SUMO targets suggested that there might be a common mechanism by

which those interaction partners bind to SUMO. Several different SUMO Interaction Motifs (SIM) have been proposed in the literature. Analysis of the sequence of the proteins already known to interact with SUMO as well as of those found by Christina Hecker led to propose a SIM sequence that reconciles the different proposals previously published. According to this analysis, the SIM consists in a hydrophobic core and a tract of negatively charged amino acids separated by one of several serine phosphorylation sites. Experimental evidence showed that the hydrophobic core of the SIM is crucial for the interaction with SUMO.

The work presented here showed what role the different elements of the SIM play in binding to SUMO. The proposal that the hydrophobic core of the SIM forms an intermolecular β-sheet with the second β-strand of SUMO has been confirmed by the structure published by [Reverter and Lima, 2005]. However, though earlier studies [Song *et al*, 2004] suggested that they might be unimportant, this work shows that the

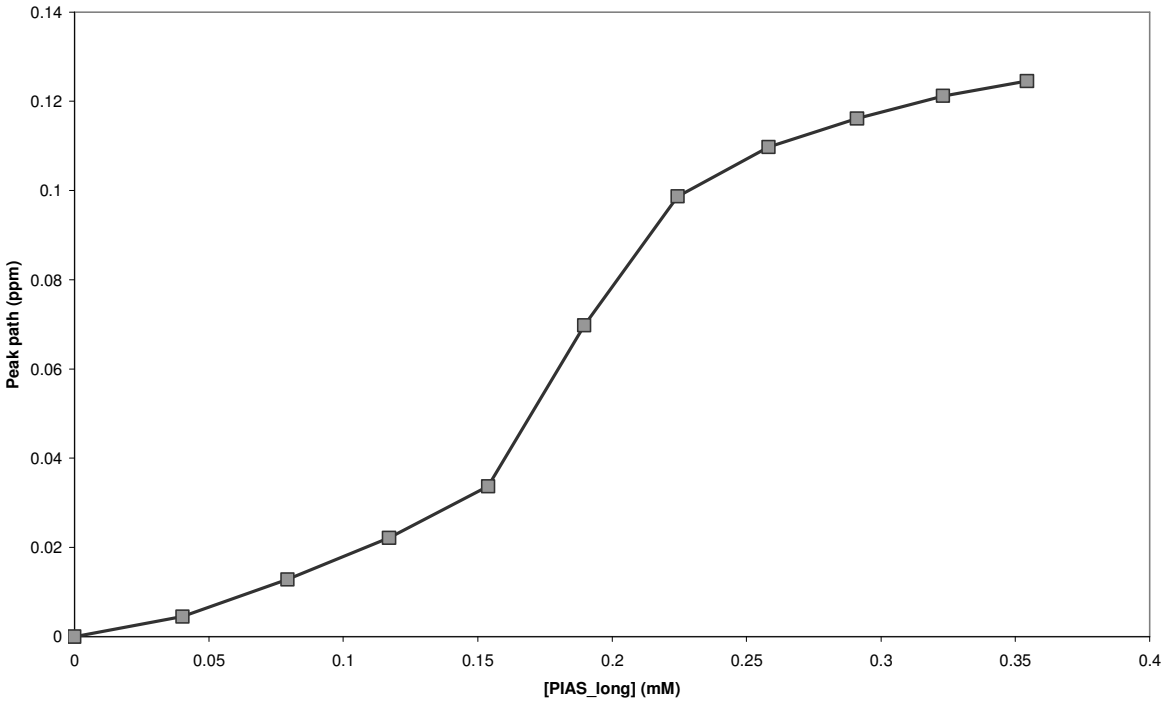


Fig. 37: Binding of PIAS_long to SUMO1 (the titration curve obtained for His 43 of SUMO1 is plotted). The peak path is proportional to the ratio $[PIAS\_long]_{free}/([PIAS\_long]_{total}$. This ratio considerably increases when the [PIAS]total increases above ~0,2 mM

amino acids flanking the hydrophobic core of the SIM play an important role the affinity of the SIM for SUMO. Thus, the SIM appears as an interaction interface that can be fine tuned by evolution. The tract of negatively charged amino acids influences the affinity of the SIM for SUMO by interacting with the loop region between the β_3 and β_4 strands of SUMO. A certain number of SIM have their negatively charges in N-term of the hydrophobic core instead of having them in C-term of it like in PIAS. This suggests that the position of those amino acids relatively to the hydrophobic core determines the orientation of the SIM relatively to SUMO. Further experiments will be done to examine this possibility. A further regulation of the affinity of the SIM for SUMO is assured by phosphorylation of serines surrounding the hydrophobic core of the SIM. In the cellular context, unlike the negatively charged amino acids, this regulation can be changed during the life of the SIM, opening the possibility of activation of SUMO interactions by phosphorylation of SIMs. In this

regard, the sigmoid curves observed for the binding of unphosphorylated PIAS to SUMO1 show that it behaves like a concentration-sensitive switch. The fraction of PIAS bound to SUMO quickly increases when a threshold PIAS concentration (~200 μM for a SUMO1 concentration of 300 μM) is reached (fig. 37). This threshold effect is absent in the binding of phosphorylated PIAS to SUMO1. At the low PIAS concentrations likely to be encountered in cells, PIAS can bind to SUMO only in its phosphorylated state. Thus the observations presented here show the structural detail of how phosphorylation “turns on” a protein for protein-protein interaction. Those variations in the SIM offer an exceptional opportunity to study the mechanisms protein-protein interactions in their finest details on biologically relevant molecules.

The interaction surface on SUMO for the SIM functions as a universal plug through which a variety of protein can be recruited. It is remarkable that Ubiquitin, to which SUMO is closely related, also has such a universal plug, the Ile 44 patch. This patch is an hydrophobic surface at the surface of Ubiquitin to which a variety of proteins bind. Despite their similar function, the SIM binding site and the Ile 44 patch are located on different areas of the ubiquitin fold. Furthermore, whereas the SIM is a well conserved short (~20 amino acids) sequence that can inserted in a protein domain, the proteins binding to the Ile 44 patch of Ubiquitin do not appear to have any common structural element for doing so. Instead, a proteins use various domains (see [Hicke *et al*, 2005] for review) to interact with the Ile 44 patch of Ubiquitin. Those differences are just another illustration of the variety of solutions selected by evolution to achieve a similar task.

References

- Baba, D., Maita, N., Jee, J. G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1, *Nature* 435: 979-982

- 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.* 280: 275-286

- Bernier-Villamor V., Sampson D.A., Matunis M.J., Lima C.D., (2002) structural basis for E2-mediated SUMO conjugation revealed by a complex between Ubiquitin-conjugating enzymes Ubc9 and RanGAP1, *Cell* 108: 345-356

- Blom N., Gammeltoft S., Brunak S. (1999) Sequence- and structure-based prediction of Eukaryotic protein phosphorylation sites, *JMB* 294: 1351-1362

- Bohren K.M., Nadkarni V., Song J.H., Gabbay K.H., Owerbach D. (2004) A M55V polymorphism of a novel SUMO gene (SUMO4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus, *J. Biol. Chem.* 279: 27233-27238

- Chupreta S., Holmstrom S., Subramanian L., Iniguez-Lluhi J.A. (2005) A small conserved surface in SUMO is the critical structural determinant of its transcriptional inhibitory properties, *Mol. Cell. Biol.* 25: 4272-4282

- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6: 277-293

- Dohmen J. (2004) SUMO protein modification, BBA 1695, 113-131

Goddard T.D. and Kneller D.G., SPARKY 3 (2004) University of California, San Fransisco [not published]

- Hannich, J. T., Lewis, A., Kroetz, M. B., Li, S. J., Heide, H., Emili, A., and Hochstrasser, M. (2005) Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae, *J. Biol. Chem.* 280: 4102-4110

- Hicke, L., Schubert, H. L., and Hill, C. P. (2005) Ubiquitin-binding domains, Nat Rev Mol Cell Biol 6: 610-621

- Huang, W. C., Ko, T. P., Li, S. S., and Wang, A. H. (2004) Crystal structures of the human SUMO-2 protein at 1.6 A and 1.2 Á resolution: implication on the functional differences of SUMO proteins, *Eur. J. Biochem.* 271: 4114-4122

- Jin C, Shiyanova T, Shen Z, Liao X. (2001) Heteronuclear Nuclear Magnetic Resonance Assignments, Structure and Dynamics of SUMO-1, a Human Ubiquitin-Like Protein *Int. J. Biol. Macromol.* 28: 227-234
- International Genome Sequencing Consortium (2001) initial sequencing and analysis of the human genome, *Nature* 409: 860-921
- Kelley L.A., Maccallum R. and Sternberg M.J.E. (1999) Recognition of Remote Protein Homologies Using Three-Dimensional Information to Generate a Position Specific Scoring Matrix in the program 3D-PSSM, *RECOMB 99, Proceedings of the Third Annual Conference on Computational Molecular Biology* 218-225
- Lapenta V., Chiurazzi P., van der Spek P.J., Pizzuti A., Hanaoka F., Brahe C. (1997) SMT3A, a novel homologue of the S. Cerevisiae MT3 gene, maps to chromosome 21qter and defines a novel gene family, *Genomics* 40: 362-367
- Lin D., Tatham M.H., Yu B., Kim S., Hay R.T., Chen Y. (2002) Identification of a substrate recognition site on Ubc9, *J. Biol. Chem.* 277: 21740-21748
- Liu, Q., Jin, C., Liao, X., Shen, Z., Chen, D. J., and Chen, Y. (1999) The binding interface between an E2 (UBC9) and a ubiquitin homologue (UBL1), *J. Biol. Chem.* 274: 16979-16987
- Lois L.M., Lima C.D. (2005) Structures of the SUMO E1 provide mechanistic insights SUMO activation and E2 recruitment to E1, *EMBO J.* 24: 439-445
- Macauley, M. S., Errington, W. J., Okon, M., Scharpf, M., Mackereth, C. D., Schulman, B. A., and McIntosh, L. P. (2004) Structural and dynamic independence of isopeptide-linked RanGAP1 and SUMO-1, *J. Biol. Chem.* 279: 49131-49137
- Mannen H., Tseng H.M., Cho C.L., Li S.S.-L. (1996) Cloning and expression of human homolog HSMT3 to yeast SMT3 suppressor of MIF2 mutations in a centromere protein gene, *BBRC* 222: 178-180
- Meluh P.B., 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.* 6: 793-807
- Mulder F.A.A., Schipper D., Bott R., Boelens R. (1999) Altered Flexibility in the Substrate-Binding Site of Native and Engineered High-Alkaline Bacillus Subtilisins, *JMB* 292: 111-123
- Pichler, A., Knipscheer, P., Oberhofer E., van Dijk W., Körner R., Olsen J.V., Jentsch S., and Melchior F. & Sixma, T. K. (2005) SUMO modification of the Ubiquitin conjugating enzyme E2-25K, *Nat. Struct. Mol. Biol.* 12: 984-991

- Reverter D., Lima C.D. (2004) A basis for SUMO protease specificity provided by the analysis of human Senp2 and Senp2-SUMO complex, *Structure* 12: 1519-1526
- Reverter, D., and Lima, C. D. (2005) Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex, *Nature* 435: 687-692
- Tatham M.H., Kim S., Yu B., Jaffray E., Song J., Zheng J., Rodriguez M.S., Hay R.T. and Chen Y. (2003) Role of an N-Terminal site for Ubc9 in SUMO1-, 2- and 3- -Binding and conjugation, *Biochemistry* 42: 9959-9969
- Tatham, M. H., Kim, S., Jaffray, E., Song, J., Chen, Y., and Hay, R. T. (2005) Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection, *Nat. Struct. Mol. Biol.* 12: 67-74
- Tong H., Hateboer G., Perakis A., Bernards R. and Sixma T.K. (1997) Crystal structure of Murine/Human Ubc9 provides insight into the variability of the Ubiquitin-conjugating system, *JBC* 272: 21381-21387

Lebenslauf

Name:	Matthias Rabiller
Anschrift:	Adlerstraβe 86, 44137 Dortmund
Geburtsdatum:	2 März 1978
Geburtsort:	Den Haag
Staatsangehörigkeit:	Französich
Familienstand:	Verheiratet

Bildungsweg

1985-1989: Grundschule
1990-1993: Collège
1994-1996: Lycée
1996: Wissenschaftliches Abitur mit Note „Assez Bien“ , Lycee Saint Cricq, Pau (Frankreich)
1996-1999: Vorbereitungsklasse für Biologie, Chemie, Physik und Geologie
1999-2002: Biotechnologisches Studium an der ESBS (Ecole Superieure de Biologie de Strasbourg, Straßburg, Frankreich)
September 2002: Biotechnologie Diplom
Oktober 2002-August 2003: Mutagenese, Klonierung, Protein Expression und Reinigung von Vitronectin Fragmenten in der Nachwuchsgruppe Molekulare und Strukturelle Biologie in Max-Planck Institut für molekulare Physiologie (Dortmund), Stipendium der Max-Planck Gesellschaft
September 2003-Oktober 2004 : Mutagenese, Klonierung, Protein Expression und Reinigung von SUMO Isoformen, Biophysikalische Studien dieser Proteine (durch CD Spektroskopie, Lichtstreuung, UV Spektroskopie, MALDI-MS Fingerprinting…), Bearbeitung von NMR Spektren an der Nachwuchsgruppe für Molekulare und Strukturelle Biologie am Max-Planck Institut für molekulare Physiologie (Dortmund), Stipendium der Max-Planck Gesellschaft
Oktober 2004 - : NMR-Studien von SUMO-Interaktionen am Zentrum für Medizinische Biotechnologie (Universität Essen) in Kooperation mit der Molecular Signaling Gruppe (Institut für Biochemie II, J.W. Universität, Frankfurt).

Praktika und Tätigkeiten

Juli-August 2000: Zweimonatiges Praktikum am Pasteur Institut (Paris), zum Thema „Analyse einer A. gambiae cDNA Bank“
Juli-August 2001: Zweimonatiges Praktikum in CRIBI (Centro Di Ricerche Interdipartimentale Biotecnologie Innovative), Universität Padua, Italien, zum Thema „Automatisierte Sammlung von Protein subzelluläre Lokalisierung Daten“
Januar-September 2002: Diplomarbeit in der Nachwuchsgruppe für Molekulare und Strukturelle Biologie, MPI für Molekulare Physiologie, mit dem Thema „Bioinformatische und biochemische Charakterisierung der Tyrosyl Protein Sulfotransferase 1“
Oktober 2003-Juli 2004: Weiterbildung im Bereich Bioinformatik an der Akademie für Weiterbildung der Universität Heidelberg (Klaus-Tschira Stipendium)

Erklärung:
Hiermit erkläre ich, gem. § 6 Abs. 2 Nr. 7 der Promotionsordnung der Fachbereiche 6 bis 9 zur Erlangung des Dr. rer. Nat., dass ich das Arbeitsgebiet, dem das Thema „Interactions of SUMO proteins“ zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Herrn Matthias Rabiller befürworte.
Essen, den 18. Januar 2006

Prof. Dr. Peter Bayer

Erklärung:
Hiermit erkläre ich, gem. § 6 Abs. 2 Nr. 6 der Promotionsordnung der Fachbereiche 6 bis 9 zur Erlangung des Dr. rer. Nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient habe.
Essen, den 18. Januar 2006

Matthias Rabiller

Hiermit erkläre ich, gem. § 6 Abs. 2 Nr. 8 der Promotionsordnung der Fachbereiche 6 bis 9 zur Erlangung des Dr. rer. Nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät abgelehnt worden ist.
Essen, den 18. Januar 2006

Matthias Rabiller