Structural Attributes in the Conjugation of Ubiquitin, SUMO and RUB to Protein Substrates

Sandra Goettsch, and Peter Bayer

Max-Planck Institute for Molecular Physiology, Research Group Molecular and Structural Biophysics, Otto-Hahn-Str.11, D-44227 Dortmund, Germany

Table of Contents

1. Abstract
2. Introduction: Post-translational modification
3. Covalent attachment of proteins
   3.1. The players: A short description
      3.1.1. Ubiquitin
      3.1.2. SUMO
      3.1.3. RUB
      3.1.4. Intertwining of pathways
   3.2. Structural Comparison of Ubiquitin, RUB and SUMO
      3.2.1. Looking at the sequence
      3.2.2. Comparison of the secondary and tertiary structures of modifiers
      3.2.3. Structure and function of E2 enzymes
4. Interaction of E2s and modifiers
5. Perspectives
   5.1. Mechanisms of target selection
   5.2. Polysumoylation?
6. Acknowledgment
7. References

1. ABSTRACT

Many cellular and secreted proteins are chemically modified after their translation is completed. The covalent linkage of a polypeptide chain (modifier) to a substrate protein is a special case of post-translational modification. In the late seventies it was observed that ubiquitin, a small modifier, marks short-lived proteins for degradation by the 26S proteasome. Over the last decade many other ubiquitin-related proteins were discovered and isolated. Attachment of polypeptide chains onto acceptor molecules became a common feature to regulate spatially and timely organized cellular pathways of proteins. This article focuses on the structures of the three modifiers: ubiquitin, RUB and SUMO and the cognate enzymes involved in these modification pathways. We have described the homologies and differences of these proteins and indicate salient topological hallmarks common to modifier-conjugating enzymes. This characterization will help in understanding these regulatory pathways and their similarities and differences in controlling protein fate, from protein degradation signals generated by polyubiquitination to functional modification brought about by RUB and SUMO conjugation.

2. INTRODUCTION: POST-TRANSLATIONAL MODIFICATION

Many cellular and secreted proteins are chemically modified after completion of their translational
Structures of protein modifiers and their conjugating enzymes

![Figure 1. Ubiquitin conjugation pathway. Ub (red) stands for ubiquitin. E1 (blue), E2 (green) and E3 (grey) mark the enzymes involved in conjugation and Sub (yellow) indicates the substrate molecule.](image_url)

process. A machinery of specific enzymes localized to different cellular compartments carries out these post-translational events. Many of these enzymes link a small chemical moiety like a phosphoryl, acetyl, sulffyl or methyl group to the side chain of a certain amino acid of an acceptor protein. A single enzyme, designated transferase or kinase, which exhibits a specific binding activity for both the donor and acceptor molecules, performs this transfer step. A special case of covalent protein modification is the attachment of a whole polypeptide chain (modifier) to a protein. In contrast to the transfer of a small moiety the linkage of a polypeptide chain is a multi-step process with different enzymes involved in its mechanism. The most prominent and historically oldest modifier known is ubiquitin. Over the last decade the discovery and isolation of a number of ubiquitin related modifiers was reported (reviewed in 1) implying that attachment of chemical moieties of preformed conjugates and catalyzes the chain assembly in conjunction with E1, E2 and E3 (28). The attachment of ubiquitin to target proteins is called ubiquitination or ubiquitin(l)ylation. It was discovered in the late seventies of the last century (6-9; reviewed in 10) and, hence, has been extensively studied. By the action of an enzyme conjugase cascade, involving a series of thiol ester formations, the C-terminal glycine of ubiquitin is linked to the ε-amino group of lysine side chains of acceptor proteins. Mono-ubiquitination, the linkage of a single ubiquitin to one or more lysine residues of a substrate molecule plays a role in receptor-mediated endocytosis or mediates virus budding and many cellular regulatory functions (reviewed in 11). The covalent attachment of a chain consisting of several ubiquitin molecules (polyubiquitin) onto one or multiple lysine residues of a protein is called multi-ubiquitination. In polyubiquitin one lysine (either Lys29, Lys48 or Lys63) of a monomer is connected to the C-terminal glycine of a neighboring unit (reviewed in 12). A multi-ubiquitination with polyubiquitin linked through Lys48 marks short-lived proteins for degradation by the 26S proteasome, the major protease in the cytosol of eukaryotic cells (13, 14). Lys63 linkages are important signals for DNA repair (15), NFκB activation (16) or ribosome function (17). Chains connected through Lys29 enhance translational activity during S-phase of cell cycling (18,19, reviewed in 20, 21). Polyubiquitin also plays a not very well defined role in neurological disorders (e.g. Alzheimer’s and Huntington’s diseases), where it is involved in the formation of intraneuronal inclusions like Lewy bodies or neurofibrillary tangles (reviewed in 22,23).

Ubiquitin is transcribed as an inactive precursor molecule with a C-terminal extension of several amino acids. A protease called ubiquitin carboxy-terminal hydrolase (UCH), which belongs to a family of papain-like thiol proteases, is involved in the maturation process of ubiquitin in yeast and higher eukaryotes (24). It processes the precursor C-terminal to the double-glycine motif. The C-terminal carboxyl group of ubiquitin is then activated by adenylation (figure 1). In a second step the thiol group of a cysteine residue of the activating enzyme E1 (e.g. Uba1 in yeast) attacks the carboxyl-AMP to form an E1-ubiquitin thiol ester bond. By trans-esterification E1 links the ubiquitin to the cysteine side chain of one of numerous conjugating enzymes (E2s or Ubc5s). E2 now passes the modifier to a substrate lysine through formation of an amide bond between the carboxy terminal carboxylate of the modifier and the ε-NH-group of the lysine side chain of the acceptor protein. For this final step, a third enzyme, the ligase E3, is required at least in most cases. Two major types of E3-ligases are known, carrying either a HECT (25) or RING (26) domain. RING domain proteins function as scaffolds to promote ubiquitination by ensuring the appropriate arrangement of E2-ubiquitin and substrate for catalysis. HECT domain proteins participate directly in catalysis by forming a thiol ester intermediate with the modifier (reviewed in 27). Multi-ubiquitination sometimes requires a fourth enzyme E4, which binds to the ubiquitin moieties of preformed conjugates and catalyzes the chain assembly in conjunction with E1, E2 and E3 (28). The chemical details of the mechanism of ubiquitination have not been studied on an atomic level so far. However, it is
3.1.4. Intertwining of pathways

Little is known about an intertwining of ubiquitin and SUMO pathways, but some observations on the activation of NFκB have demonstrated that these pathways exist (68, reviewed in 1). The inhibitor of NFκB, IκB, is regulated and degraded by ubiquitination. From mutagenesis studies it has been inferred that a lysine residue of IκB is modified by both SUMO and ubiquitin. RUB and SUMO pathways are directly connected. Cullins, the targets of rubylation, are part of SCF or CBC complexes, two large multimer units, which function as ubiquitin-ligases. The Nedd8 modifying-system can accelerate the formation of the E2-E3 complex and, thereby, stimulate protein polyubiquitination (60). It was shown, that conjugation of Nedd8 to ROC1-CUL1, a subcomplex of the SCF-ROC1 E3 ubiquitin ligase, selectively stimulates Cdc34-catalyzed lysine 48-linked multoubiquitin chain assembly (69).

3.2. Structural Comparison of Ubiquitin, Rub and Sumo

3.2.1. Looking at the sequence

A detailed sequence comparison of human ubiquitin with RUB and SUMO from different organisms and a scheme highlighting residues involved in binding to E2 enzymes are shown in figure 2. Red colors denote amino acids that are strictly conserved from human to yeast in all three modifiers. These residues in ubiquitin are Val5<sup>ub</sup>, Gly<sup>75ub</sup>, Thr<sup>55ub</sup>, Gly<sup>76ub</sup> and Gly<sup>76ub</sup>. The last two glycines are important for conjugation and reversible proteolytic deconjugation (61, 62; reviewed in 30). Gly<sup>76ub</sup> to Ala mutants result in derivatives that become irreversibly attached to substrate molecules, whereas deletions in
Structures of protein modifiers and their conjugating enzymes

Gly75\textsuperscript{ub} / Gly76\textsuperscript{ub} prevent growth of modified yeast strains (63). In the same experimental approach Gly47\textsuperscript{ub} in ubiquitin was shown to be essential for vegetative growth in yeast. The function of this residue is not clear, yet. It is often suggested that the importance of Gly47\textsuperscript{ub} is due to its direct neighborhood to Lys48\textsuperscript{ub}. However, this seems to be unlikely considering the fact, that Gly68\textsuperscript{SUMO} is conserved at the respective sequence position in SUMO, although Gln69\textsuperscript{SUMO} does not play any role in conjugation. From figure 2 it is obvious that the conserved Gly47\textsuperscript{ub}/68\textsuperscript{SUMO}, which is part of a small loop connecting \(\beta_3\) and \(\beta_4\), is located within the modifier/E2-binding interface. Thus, it might be important for the interaction with the corresponding Ubc protein. Val5 is a central part of the hydrophobic core of the modifier, which probably explains the conservation of this residue. Sequence comparison alone gives no information on the conservation needs for Thr55.

Some amino acids in figure 2 exhibit either a high (blue color) or medium (yellow color) degree of conservation through all groups of modifiers and all species. These residues are mainly involved in the hydrophobic core or in salt bridges and, hence, are important for the structural integrity of the protein. The most interesting amino acids are either the green and

### Figure 2. Sequence comparison of modifiers. Sequence numbers indicate the corresponding residues in human ubiquitin (ubiq).

RUB-related proteins are abbreviated by ned8 and SUMO-related modifiers by sumo or smt3. Sequences were derived from the SWISS-PROT/TrEMBL databank or GenBank (accession codes: P02248, Q03968, O57686, DaQ9SMD1, Q12306, Q15843, P29595, NP_010423, AAC17623). Amino acids are classified by color code: red, identical in all modifiers; blue, homology in all groups of modifiers >75%; yellow, homology in all modifiers >40% and <75%; black, no homology or minor homology; green, homology in the RUB/ubiquitin group >80%; margenta, homology in the SUMO group >75%. Modifier residues involved in binding to enzymes or residues that show chemical shift changes in HSQC-spectra on complex formation are marked by x. The complexes used are ubiquitin/Ubc2b (u-ubc2b), ubiquitin/Ubc1(u-ubc1), thiol ester-linked ubiquitin/Ubc1(u-ubc1t), ubiquitin/Ulp1(u-ulp1) and SUMO-1/Ubc9(s-ubc9).

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ1FVKTGKTITLEVPSVTENVKAKIQ ubiq hs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML1KVKTGKEIEIDPTQKIERVE ned8 hs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML1KVKTGKEIEIDPTQKIERVE ned8 mu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIV1KVKTGKEIEIDPTQKIERVE ned8 sc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME1KVKTGKEIEIDPTQKIERVE ned8 at</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2: Sequence comparison of modifiers.** Sequence numbers indicate the corresponding residues in human ubiquitin (ubiq). RUB-related proteins are abbreviated by ned8 and SUMO-related modifiers by sumo or smt3. Sequences were derived from the SWISS-PROT/TrEMBL databank or GenBank (accession codes: P02248, Q03968, O57686, DaQ9SMD1, Q12306, Q15843, P29595, NP_010423, AAC17623). Amino acids are classified by color code: red, identical in all modifiers; blue, homology in all groups of modifiers >75%; yellow, homology in all modifiers >40% and <75%; black, no homology or minor homology; green, homology in the RUB/ubiquitin group >80%; margenta, homology in the SUMO group >75%. Modifier residues involved in binding to enzymes or residues that show chemical shift changes in HSQC-spectra on complex formation are marked by x. The complexes used are ubiquitin/Ubc2b (u-ubc2b), ubiquitin/Ubc1(u-ubc1), thiol ester-linked ubiquitin/Ubc1(u-ubc1t), ubiquitin/Ulp1(u-ulp1) and SUMO-1/Ubc9(s-ubc9). Secondary structure elements (sec) are marked by h(low probability)/H(high probability) for helices (red) or b(low probability)/B(high probability) for β-strands (green).
Structures of protein modifiers and their conjugating enzymes

![Figure 3](image-url)

**Figure 3.** Molscript representations of the modifiers: a, ubiquitin (RCSB entry 1ubq), b, Nedd8 (1ndd), c, SUMO-1 (1a5r). The N- and C-termini of the proteins are labeled by capital letters.

magenta colored residues, since they determine the difference between the SUMO and the ubiquitin/RUB group of modifiers, or the black colored amino acids, since they might provide the origin of individuality (binding specificity and selectivity for enzymes). This statement is especially true for amino acids located in loop and turn regions. One very good example for this statement, which has been extensively studied is the role of amino acid 72 in RUB and ubiquitin. An Arg72 in Ala mutation was shown to alter the binding of ubiquitin-adenylate to its E1 (64) and an Ala72 in Arg72 mutation in Nedd8 bound to the ubiquitin directed E1 almost as well as wild-type ubiquitin (65). Residue Ala72 in RUB is thought to perform a key role in selecting against reaction with the ubiquitin-specific E1 enzyme and, thereby acting to prevent the inappropriate diversion of RUB into ubiquitin-specific pathways (65).

On basis of the sequence alignment the SUMO and ubiquitin/RUB families are very well separated from each other. One could hypothesize, that only few amino acids are determinants for the group character. For example the green (magenta in sumo) colored residues (figure 2) Thr7, Lys11, Glu/Asp16, Glu/Tyr21, Glu24/30, Arg33/35, Pro38/40, Gln40/49, Gln41, Arg42/43, Ile44, Gln49, Arg50, His56, Leu73/74, Leu73/74, and Arg74/75 separate the ubiquitin/RUB family from the SUMO group. Clearly, the polarity (charges) of side chains at positions 9, 11, 16, 21, 24, 40, 49 and 74 of ubiquitin/RUB is changed in SUMO.

Recent studies on ubiquitin (66,67) have revealed residues Phe4 and Ile44 to be crucial for internalization and Leu8, Ile44 and Val70 for proteasome binding and degredation of proteins. Phe4 is unique for ubiquitin, whereas RUB and SUMO proteins carry a lysine at this position. This lysine can conserve part of the hydrophobic character of Phe4, but at the same time introduces a positive charge close to the hydrophobic patch surrounding this residue. It might prevent RUB and SUMO to compete with ubiquitin for internalization. Additionally, there are some minor differences that separate the SUMO class from the ubiquitin/RUB group. For example, sequence position 44 is occupied by isoleucine or valine in ubiquitin and RUB, but is substituted by leucine in SUMO. Also Leu8 and Val70 are replaced by various other residues in SUMO. The observation that the hydrophobic residues Leu8, Ile44 and Val70, which are crucial for proteasome binding, are strictly conserved among the ubiquitin/RUB family, but not within the SUMO group, kicks SUMO out of the degradation pathway, however, at the same time leaves a door open for RUB.

### 3.2.2. Comparison of the secondary and tertiary structures of modifiers

Ubiquitin, RUB and SUMO consist each of a five-stranded β-sheet with antiparallel aligned single strands, an α-helix connecting β2 and β3 and a short helix between β4 and β5 (figure 3). This arrangement represents the typical ubiquitin superfold (70). There are many variations on this theme and the ubiquitin superfold has become a paradigm for structural conservation in absence of significant sequence identity. It is found in proteins that are functional entirely unrelated to ubiquitin like the (2F-2S)-ferredoxin (71), the Ras-binding domains (72-74) and numerous other proteins.

The topologies of AtRUB1 and ubiquitin are very similar. The root mean square deviation (rmsd) of the two sets of α-carbon atoms is 0.612 Å (54) or 0.8 Å (DALI search). The rmsd value between SUMO-1 and ubiquitin (75) and SUMO-1 and Nedd8 (54) are 2.1 Å and 2.6 Å, respectively. The hydrophobic core of ubiquitin/RUB (SUMO) is made by side chains of residues 3α/3α, 4α/4α, 13α/13α, 16α/16α, 24α/24α, 44α/44α, 65α/65α, 67α/67α for ubiquitin and SUMO, respectively. The hydrophobic core of ubiquitin/RUB (SUMO) is made by side chains of residues 3α/3α, 13α/13α, 16α/16α, 24α/24α, 44α/44α, 65α/65α, 67α/67α for ubiquitin and SUMO, respectively.

Several salt bridges are stabilizing the topology of the modifiers. In the structures of ubiquitin and AtRUB1 Lys27/27, Arg33/33, Lys34/34 and Asp52/52 (2.74 Å and 2.90 Å) form a salt bridge and a second one exists between Lys11/11 and Glu34/34 (54.3 Â) (54). In SUMO His75 and Glu79 are involved in a salt bridge (2.6 Å) (75).

Structure comparison allows to explain the conservation of Thr55 in all three modifier groups, although mutations in this position had no effect on the viability of yeast cells (63). Thr55 is located at the surface of the protein and a detailed analysis of the structures of ubiquitin and Nedd8 reveals the side chain oxygen group to form a hydrogen bond to the amino group of Asp76 in RUB, thereby stabilizing helix2. Unfortunately, the SUMO structure is only weakly determined in this region (75).

A major structural difference between the ubiquitin/RUB class of proteins and its SUMO relatives is found at the N-termini of the modifiers. SUMO-1 exhibits a flexible and unstructured 21 residue long extension preceding the ubiquitin-like core domain. The function of this tail remains unclear. Homology searches (FASTA3) using the N-terminus as bait revealed 50% identity to...
structures of protein modifiers and their conjugating enzymes

sequences of HMG1 protein and chicken filamin. Ser9 and Tyr21 were detected as possible phosphorylation sites in SUMO-1 using the web engine of Blom (76). However, no post-translational modifications have been observed in SUMO until now.

3.2.3. Structure and function of E2 enzymes

E2 enzymes (Ubcs) are predominantly localized to the nucleus (77). They can be classified into four structural categories: enzymes that are comprised of a core domain (about 150 residues) carrying the active site cysteine (class I), those who additionally have either a C- or N-terminal extension (class II and III), and a group of proteins with both C- and N-terminal tails (class IV). Ubcs or N-terminal extension (class II and III), and a group of cysteine (class I), those who additionally have either a C-

Non-conserved residues in the E2-modifier complex are the major determinants for selection of the proper ligase. Amino acids of Ubc13 critical for Mms2 binding are Tyr34ubc13, Phe57ubc13 and Leu83ubc13. In the case of the E6Ap(HECT)-Ubc7h complex Phe63ubc7h is the primary determinant for E2 specificity. Together with Ala59ubc7h, Glu60ubc7h Pro62ubc7h and the amino acids Lys96ubc7h, Ala98ubc7h and Lys100ubc7h the residue Phe63ubc7h is involved in direct contacts to the HECT domain (80). The fact that non-conserved residues are the origin of individuality of proteins also holds for other Ubcs. For example, deletion of the N-terminal nine amino acids (part of the first helix) of Rad6 (Ubc2sc) prevents the enzyme from forming a stable complex with the E3-ligase Ubr1 (81). The same was observed upon deletion of the solvent exposed residues Glu150Rad6, Asp151Rad6, Asp152Rad6 and Met153Rad6, which are part of the C-terminus of Rad6.

4. INTERACTION OF E2s AND MODIFIERS

NMR titration experiments provide the necessary platform for a detailed comparative view on complex formation of Ubcs and their respective modifiers. Results of shift perturbation studies allowed mapping of the binding interfaces of ubiquitin and the two E2 conjugating enzymes Ubc1 and Ubc2b (82-84). HSQC spectra of free ubiquitin revealed changes in 1H and 15N chemical shifts of numerous resonances on the addition of Ubc2h. Residues belonging to these resonances cluster on one side of the protein surface (figure 6A) around the C-terminal Gly-Gly motif (red color). A collection of basic amino acids (Arg42ub, Lys48ub, Arg72ub, Arg74ub) is present in this region. Thus, the binding interface of ubiquitin is positively charged (figure 6B). The side chains of residues Leu8ub, Ile44ub, Val70ub, Leu71ub and Leu73ub provide hydrophobic patches for the interaction with Ubc2b. Thr76ub, Thr9ub, Gln40ub and Gln49ub can serve as hydrogen bond donors or acceptors. Similar results could be obtained by shift perturbation experiments using an Ubc1-modifier complex linked by a thiol ester bond. The binding interface presented in figure 6C/D was mapped on the surface of ubiquitin (green color) after computational docking based on NMR data. In figure 6C/D the C-terminus of the modifier sticks out of the paper plane. The same amino acids
Structures of protein modifiers and their conjugating enzymes

Figure 5. Sequence comparison of human (h) and yeast (sc) E2 enzymes. Sequence numbers indicate the corresponding residues in Ubc4 of Saccharomyces cerevisiae. Ubc enzymes used in this context are indicated by there Ubc number and the respective species abbreviation, e.g.12sc stands for Ubc12 of yeast. Sequences were derived from the SWISS-PROT/TrEMBL databank (accession codes: P15731, P51669, Q9Y2X8, P51966, Q02159, P23567, P06104, Q76689, P52491, P50550, P50623, P31734, Q16781, P52490). Amino acids are classified by color code: red, identical in all E2s; blue, homology in all groups of E2s >75%; yellow, homology in all E2s <75% and >50%; black, no homology or minor homology; green, homology in Ubc12 group >75%; magenta, homology in the SUMO group >75%. Residues of Ubcs involved in binding to modifiers, residues of Ubc13 known to contact mms2 and residues that show chemical shift changes in HSQC-spectra on complex formation are marked by x. The complexes used for interface identification are ubiquitin/Ubc2b (i2h), ubiquitin/Ubc1 (i1sc), SUMO-1/Ubc9 (i9h) and Ubc13/Mms2 (mms2). Secondary structure elements (sec) are marked by h(low probability)/H(high probability) for helices (red) or b(low probability)/B(high probability) for β-strands (green). A consensus sequence of E2 (ubiquitin/RUB group) is indicated by ubR and for the SUMO group by sum. Grey color indicates residues that differ between SUMO and ubiquitin/Nedd8 family.
Structures of protein modifiers and their conjugating enzymes

Figure 6. Binding interface and surface charges of ubiquitin. a, Molecular surface representation (GRASP) of ubiquitin. Residues, which show chemical shift changes upon addition of Ubc2b are mapped onto the surface by red color. b, Surface charges across the binding interface. Negative charges are mapped in red, positive ones in blue. c, Molecular surface representation (GRASP) of ubiquitin complexed to Ubc1. Residues, which are involved in direct contacts to the thiol ester-linked Ubc1 are colored in green. d, Surface charges of complexed ubiquitin. Negative charges are colored in red, positive ones in blue.

Figure 7. Comparison of the surface charges of the binding interfaces of a, ubiquitin and b, Nedd8. Negative charges are colored in red, positive ones in blue.

acids, which are involved in contacts to Ubc2b, form the modifier interface pointing towards Ubc1. These residues are mainly located in the β-sheet and the loop region connecting α3 and β5 of ubiquitin.

If ubiquitin always presents the same interface towards an E2 enzyme and only one E1 enzyme is used for activation, how can the E1-ubiquitin complex select for different Ubcs? One can currently think of two ways on how to answer this question. First, the “proper” Ubc is chosen by no selection criteria and each Ubc has the same likelihood to be accepted. This would mean, that probability of complex formation is proportional to the amount of Ubcs present in the cell. Selection would then be achieved by regulating the expression of E2 enzymes. Second, the affinities of Ubcs to the E1-ubiquitin complex are different, thereby, serving as a selection mechanism. Probably, the cell will use both criteria at the same time.

RUB closely resembles the surface properties of ubiquitin (figure 7). If one defines the binding interface of RUB on the basis of the structural homology of both
Structures of protein modifiers and their conjugating enzymes

Figure 8. Binding interface and surface charges of human SUMO-1. a. Surface charges within the binding interface. Negative charges are colored in red, positive ones in blue. b. Molecular surface representation (GRASP) of SUMO-1. Residues, which show chemical shift changes upon addition of Ubc9 are marked by cyan color.

proteins, nearly identical hydrophobic patches and charges are placed at comparable positions. Taking only homology criteria into consideration, RUBs might be able to use the ubiquitin E2-machinery. Thus, discrimination between the two pathways must occur at an earlier stage of the conjugase cascade, e.g. at the E1 branch point (see Ala72 of RUB).

The binding interface of SUMO is built up analogous to those of RUB and ubiquitin and is formed by residues located in the β-sheet and the neighboring loop regions. In contrast to RUB and ubiquitin, SUMO carries an extremely negative surface charge (figure 8 A and B). The side chains of Glu33\textsuperscript{SUMO}, Glu67\textsuperscript{SUMO}, Glu83\textsuperscript{SUMO}, Glu85\textsuperscript{SUMO}, Glu93\textsuperscript{SUMO} and Asp86\textsuperscript{SUMO} are directly involved in or positioned close to the interface. On the basis of NMR shift perturbation experiments no decision could be made on the functional role of Lys25\textsuperscript{SUMO}, the most prominent basic residue, which is conserved in different species (85). The same problem was faced for Arg63\textsuperscript{SUMO} and Arg70\textsuperscript{SUMO}, the other two positive charged amino acids flanking the interface.

Currently, there is no solution or crystal structure of any modifier-E2 complex available in the databank, but, as already mentioned, an ubiquitin-Ubc1 model derived from computational docking and NMR shift perturbation studies (83). Relying on this model one can find those amino acids that make direct contacts between both molecules. In figure 9 the corresponding interface is mapped on Ubc1 (figure 9B) and surface charges across the E2 binding site are marked (figure 9A). The interface area is about 1200 Å\(^2\) and consists of a large hydrophobic patch formed by amino acids Leu89\textsuperscript{Ubc1}, Ile91\textsuperscript{Ubc1}, Leu92\textsuperscript{Ubc1}, Trp96\textsuperscript{Ubc1}, Val99\textsuperscript{Ubc1}, Ile100\textsuperscript{Ubc1}, Ala105\textsuperscript{Ubc1}, Ile107\textsuperscript{Ubc1} and Leu112\textsuperscript{Ubc1}, the two negative charged residues Glu117\textsuperscript{Ubc1} and Asp120\textsuperscript{Ubc1} and the hydrogen bond donors (acceptors) Asn79\textsuperscript{Ubc1}, Ser108\textsuperscript{Ubc1}, Gln114\textsuperscript{Ubc1}, Ser115\textsuperscript{Ubc1} and Asn119\textsuperscript{Ubc1}. The side chains of Leu89\textsuperscript{Ubc1} and Leu112\textsuperscript{Ubc1} make van der Waals-contacts to the \(\gamma\)-protons of Val70\textsuperscript{ub} and Ile91\textsuperscript{Ubc1}.
null
Structures of protein modifiers and their conjugating enzymes

Table 1. Topological conserved residues

<table>
<thead>
<tr>
<th>Ubc1sc</th>
<th>Ubc2sc</th>
<th>Ubc4sc</th>
<th>Ubc7sc</th>
<th>Ubc9h</th>
<th>Ubc12h</th>
</tr>
</thead>
<tbody>
<tr>
<td>K103</td>
<td>R15?</td>
<td>K102</td>
<td>K118</td>
<td>K110</td>
<td>Part missing</td>
</tr>
<tr>
<td>E117</td>
<td>D115</td>
<td>D113</td>
<td>E129</td>
<td>E122</td>
<td>E142</td>
</tr>
<tr>
<td>S108</td>
<td>S108</td>
<td>S106</td>
<td>S122</td>
<td>?</td>
<td>S131</td>
</tr>
<tr>
<td>A111</td>
<td>S111</td>
<td>S109</td>
<td>S125</td>
<td>F22?</td>
<td>F138?</td>
</tr>
<tr>
<td>I107</td>
<td>T107</td>
<td>L105</td>
<td>L121</td>
<td>L114</td>
<td>T134</td>
</tr>
<tr>
<td>L112</td>
<td>L112</td>
<td>L110</td>
<td>M126</td>
<td>L119</td>
<td>L139</td>
</tr>
<tr>
<td>L89</td>
<td>L89</td>
<td>L87</td>
<td>I90</td>
<td>L94</td>
<td>L116</td>
</tr>
<tr>
<td>I91</td>
<td>I91</td>
<td>I89</td>
<td>I92</td>
<td>I96</td>
<td>I118</td>
</tr>
<tr>
<td>I87</td>
<td>I87</td>
<td>I85</td>
<td>V88</td>
<td>V92</td>
<td>V114</td>
</tr>
<tr>
<td>N119</td>
<td>N117</td>
<td>N115</td>
<td>N131</td>
<td>N124</td>
<td>N144</td>
</tr>
<tr>
<td>C88</td>
<td>C88</td>
<td>C86</td>
<td>C89</td>
<td>C93</td>
<td>C115</td>
</tr>
</tbody>
</table>

1 Residues that are conserved at identical spatial position upon all E2 surfaces are indicated. The headline of the table indicates the corresponding E2 enzyme. Question marks are used for those positions that could not be assigned unambiguously to a certain residue.

Figure 12. Binding interface and surface charges of Ubc9. a and b, Surface charges within the binding interface. Negative charges are colored in red, positive ones in blue. c and d, Molecular surface representation (GRASP) of Ubc9. Residues, which showed chemical shift changes upon addition of Ubc9 are colored in magenta. b and d are rotated with respect to A by 90° as marked by the arrow.

where it seemed to be sufficient to use non-covalently linked modifier to map specific residues within the interface of Ubc1. In the case of Ubc1/2 complex formation without the thiol ester bond no crucial differences in the interface areas to the thiol ester conjugated complex were observed (82,84).

5. PERSPECTIVES

5.1. Mechanisms of target selection

Different substrate molecules are specifically recognized and degraded by the ubiquitination machinery due to the combined action of different Ubcs and substrate specific E3-ligases. The acceptor sites for ubiquitination of target molecules send to the proteasome are quite different and depend on the recognition site of the corresponding E3 ligase. These motifs are called degrons and each degron has a corresponding E3 ligase.

In contrast to ubiquitin and RUB conjugation targets it was previously suggested that PEST sequences within the substrate might be necessary for the SUMO directed enzyme machinery to choose the right domain for sumoylation. However, there are many potential and known targets without carrying such a motif (41). Acceptor sites for sumoylation of proteins contain the single consensus motif aKxE (a is a hydrophobic amino acid and x can be anyone) that seems to be present in the majority of target molecules (41), but is not strictly conserved in all known substrate molecules (table 2).

So, how does the E2-E3-SUMO complex select for target proteins? Does it really grip a four amino acid motif? No structure of the E3 ligase or a sumoylated protein is known and thus one can only speculate on how sumoylation sites are recognized. According to a Prosite search 70% of all eukaryotic sequences carry at least one aKxE pattern, hence, all of these proteins would be potential targets for SUMO. As far as we know, this seems to be unlikely and indicates an underestimation of the actual size of the recognition motif. There are probably other aspects that have to be taken into consideration, for instance, the influence of secondary or tertiary structure of the target molecule. The motif could reside in a special element (α-helix, β-sheet etc.) or be located in close proximity to a topological feature of the protein (hydrophobic patch, charge, etc). It can either be recognized by just one of the binding partners of the E2-E3-SUMO complex or by an interface formed by two or three molecules of the multimer. To present a possible scenario what one might have to look for, the recognition sites of numerous proteins, known to be sumoylated (41), were examined for secondary structure elements. GoriV and nnPredict was used for the secondary structure prediction. Based on the algorithm the sequences can be separated into two classes. The first class contains α-helical structures and the second class contains the C- and N- termini of proteins or loop regions, which
Structures of protein modifiers and their conjugating enzymes

Table 2. Sumoylation sites

<table>
<thead>
<tr>
<th>Number</th>
<th>Acceptor protein</th>
<th>Residue number</th>
<th>Sequence pattern</th>
<th>Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PML</td>
<td>K490</td>
<td>PRKVIMESSE</td>
<td>helix</td>
</tr>
<tr>
<td>2</td>
<td>RanGAP1</td>
<td>K526</td>
<td>HMGLLKEVDK</td>
<td>helix</td>
</tr>
<tr>
<td>3</td>
<td>PML</td>
<td>K160</td>
<td>HQWFLKEARP</td>
<td>helix</td>
</tr>
<tr>
<td>4</td>
<td>IkBa</td>
<td>K21</td>
<td>PRDGLKKERL</td>
<td>helix</td>
</tr>
<tr>
<td>5</td>
<td>p53</td>
<td>K386</td>
<td>KKLKFTEGPD</td>
<td>helix-coil</td>
</tr>
<tr>
<td>6</td>
<td>PML</td>
<td>K65</td>
<td>CQAEEAKCPKL</td>
<td>helix-loop-helix</td>
</tr>
<tr>
<td>7</td>
<td>Cdc3</td>
<td>K4</td>
<td>MSLKEEQVS</td>
<td>N-terminus-helix</td>
</tr>
<tr>
<td>8</td>
<td>Cdc3</td>
<td>K11</td>
<td>EQVSIKQDIQ</td>
<td>helix-loop-helix</td>
</tr>
<tr>
<td>9</td>
<td>c-Jun</td>
<td>K229</td>
<td>RLQALKEEPQT</td>
<td>β-loop-β</td>
</tr>
<tr>
<td>10</td>
<td>Mdm2</td>
<td>K446</td>
<td>CQPRPKNGCV</td>
<td>β-loop-β</td>
</tr>
</tbody>
</table>

Source: Ref 41. The sequences are numbered from 1 to 10 in column one. Second column, Names of target molecules. Third column, Number of acceptor lysine residue in the sequence. Fourth column, Sequence pattern containing the acceptor lysine. Fifth column, Secondary structure prediction of sequence. Italic letters indicate flanking sequences.

Figure 13. Helix-wheel representation of amino acid sequences. a, Sequences predicted to be helical (table 2) were ordered in circles around a helix-wheel. Green color marks hydrophobic residues, red color acidic and blue color basic ones. b, 3D-representation of a. Residues are represented by sticks.

connect α-β or β-β elements. The aKXE motif is less conserved in the latter class of sequences. The position of the aKXE motif within extended helices is close to the C-terminal end. Hence, the recognition pattern prefers to be located at an exposed site of the molecule.

In figure 13 the sequences of table 2 are shown in a helix-wheel representation. The side chains of the conserved lysine and glutamic acid residues are oriented towards opposite sides. Mainly hydrophobic residues surround the glutamic acid, whereas the helix side carrying the lysine is preferentially polar. At least one additional positively charged group is present in either position 2 or 10 of the wheel.

Although the detailed mechanism of modifier attachment onto protein targets is not known, one can suggest that the substrate molecule has to come close to the active site cysteine of Ubc9. As an example, the hydrophobic groove on the E2 surface, which is not saturated by SUMO, could pick up the hydrophobic part of the helix. The conserved glutamic acid might form a salt bridge to Arg104 and the positive charge on the opposite side could face Glu118 or Glu122, thereby stabilizing a conformation, wherein Lys could be linked to the target protein. The aKXE motif of loop regions might be able to adopt a similar conformation by an induced fit mechanism. But, of course, this is just one possible hypothesis.

5.2. Polysumoylation?

It is often stated that the reason for the absence of polysumo chains is the substitution of three lysines in SUMO, which were shown to be key residues for polyubiquitin formation. However, there are no obstacles for using any other lysine for synthesis of polysumo chains. One such candidate is Lys39, which is on the opposite side of the SUMO binding interface. Also lysine residues in the flexible N-terminus might be possible target sites.

Recent studies on the structure of Ubc13-Mms2 (78), the complex catalyzing the formation of Lys63-linked ubiquitin polymers, have shed some light onto the mechanism of polyubiquitin synthesis. Mms2 is an ubiquitin E2 variant (UEV) lacking the cystein that is necessary for thiol ester formation in Ubc13. The heterodimer forms a T-shaped complex, which enables the linkage of two ubiquitin monomers head-to-tail. The efficiency of polyubiquitin synthesis is increased by binding of Rad5 (homologue of Traf6), an E3 ligase belonging to the RING domain family of proteins.

The question is now: is there a SUMO E2 variant that can form complexes with Ubc9 or a non-discovered Ubc? The answer is not straightforward, but there are some interesting findings recently published in a paper on the SUMO E3 ligase PIASy (52). Immunoblot (SDS-PAGE) studies were performed to demonstrate the activity of the enzyme by sumoylation of the wnt-dependent transcription factor Lef1. In some of the figures SUMO-Lef1 conjugation products are shown. Surprisingly, there are ladders of several different species of Lef1-SUMO conjugates visible. As the blots do not have any molecular weight markers and Lef possesses two putative sumoylation sites, one can only speculate, whether these blots might indicate the presence of multi-sumoylation.

6. ACKNOWLEDGEMENT

We like to thank Jonathan Mueller and Elena Bayer for carefully reading the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (BA 1624/3-2 and 1624/4-1) and the Max-Planck Institute for Molecular Physiology Dortmund. We apologize to many authors that could not be cited due to space constraints.
Structures of protein modifiers and their conjugating enzymes

7. REFERENCES


36. Matunis M. J, E. Coutavas & G. Blobel: A novel ubiquitin-like modification modulates the partitioning of

Send correspondence to: Dr. Peter Bayer, Max-Planck Institute for Molecular Physiology, Head of Research Group, Molecular and Structural Biophysics, Otto Hahn Straße 11, D-44227 Dortmund, Germany, Tel.: ++49-231-133-2222, Fax: ++49-231-133-2699, E-mail: peter.bayer@mpi-dortmund.mpg.de