学位論文

A screening method for identifying SUMOylated proteins based on reconstitution of split fluorescence protein fragments

(二分割蛍光タンパク質の再構成に基づく

SUMO 化タンパク質同定のためのスクリーニング法)

平成 28 年 12 月博士（理学）申請

東京大学大学院理学系研究科

化学専攻

小宮　麻希

学位論文

A screening method for identifying SUMOylated proteins based on reconstitution of split fluorescence protein fragments

(二分割蛍光タンパク質の再構成に基づく

SUMO 化タンパク質同定のためのスクリーニング法)

平成 28 年 12 月博士（理学）申請

東京大学大学院理学系研究科

化学専攻

小宮　麻希

Acknowledgments

I thank Prof. Takeaki Ozawa for constant support.

I also thank members of the Ozawa laboratory, especially Dr. Mitsuru Hattori, Dr. Mizuki Endo, and Mr. Daisuke Hiruma for helpful instructions, suggestions and discussions.

My thanks as well go to members of the Yoshida laboratory (The Chemical Genetics Lab. at Riken), especially Dr. Minoru Yoshida, Dr. Akihiro Ito, and Ms. Mikako Hirohama for assistance in analysis of *in vitro* SUMOylation of Atac2 and helpful suggestions.

I am also grateful to Dr. Hisato Saito (Department of New Frontier Sciences, Graduate School of Science and Technology, Kumamoto University) for kindly giving me purified recombinant His-SUMO2 proteins.

Abstract

Small Ubiquitin-related Modifier (SUMO) proteins are post-translational protein modifiers, found in all eukaryotes. It reversibly attaches to the lysine residues in substrate proteins and variously alters biological functions of modified proteins. SUMOylation is related to various cellular events, thereby playing essential roles in sustaining cell activities. The discovery of unreported SUMOylated proteins is prospective for understanding of the SUMOylation regulatory mechanisms in cells. So far several screening methods for SUMOylated proteins have been devised, such as immunoprecipitation-based MS analytical method and yeast two-hybrid method. However, the reversible and low-level modification of SUMOylation still makes it difficult to detect mammalian SUMOylation. For screening infrequently-SUMOylated mammalian proteins, a new method is required. **(Described in Chapter 1)**

Here, I established a new screening method for detecting mammalian SUMOylated proteins. The present method was based on the reconstitution of split fluorescence protein fragments and a cell sorting by a fluorescence-activated cell sorter (FACS). The fluorescence protein reconstitution enabled to detect protein-SUMO interaction under a live-cell condition as fluorescence signals and maintain the signals due to the irreversible reconstitution reaction. Based on the reconstitution technique, SUMOylation of various proteins that were coded by mouse cDNA library was assessed in individual mammalian cells. The cells emitting fluorescence due to the fluorescence protein reconstitution induced by SUMOylation of library proteins were selectively collected by FACS. DNA analysis of the collected fluorescent cells resulted in detection of 38 mouse-derived SUMOylated protein candidates modified by SUMO2, a mammalian SUMO family member. The detected proteinsreportedly showed different subcellular localization and functions. Among the candidate proteins, two proteins were already reported to be SUMOylated by SUMO2 in mammalian

cells. This ensured that the present system using the reconstitution of split fluorescence protein fragments and FACS sorting enabled to detect SUMOylated proteins. **(Described in Chapter 2)**

To identify novel SUMOylated proteins, SUMOylation of the candidate proteins that have not been reported as SUMOylated, was examined by immunoprecipitation and Western blotting analysis. As a result, Atac2, a histone acetyltransferase, was confirmed to be SUMOylated in mammalian cells. In addition, further analyses were conducted to reveal the characteristics of Atac2 SUMOylation, such as identification of the SUMO-binding site in Atac2, examination of Atac2 SUMOylation under *in vitro* condition, and the subcellular location of SUMOylated Atac2 through microscopic analysis. Though the cellular function of Atac2 SUMOylation is still remained to be examined, clarification of the biological effect of Atac2 SUMOylation will provide a new insight into SUMOylation roles. **(Described in Chapter 3)**

Table of Contents

Abbreviation

Rps9	Ribosomal protein S9
Rpl37a	Ribosomal protein L37a
Eif3e	Eukaryotic translation initiation factor 3, subunit E
Lmna	Lamin A
Gsn	Gelsolin
Stx12	Syntaxin 12
Uqcrh	Ubiquinol-cytochrome c reductase hinge protein
Tmsb4x	Thymosin, beta 4, X chromosome
Ermp1	Endoplasmic reticulum metallopeptidase 1
Plxnb2	Plexin B2
Rpl18a	Ribosomal protein L18A
Mrpl4	Mitochondrial ribosomal protein L4
Atac2	ADA2A-containing complex subunit 2
Rpsa	ribosomal protein SA
Lgals3	lectin, galactose binding, soluble 3
Zfp862	zinc finger protein 862
Pcolce	procollagen C-endopeptidase enhancer protein
Tuba1b	tubulin, alpha 1B
Bgn	Biglycan
Pbrm1	polybromo 1
Myof	Myoferlin
Dynlrb1	dynein light chain roadblock-type 1
Fam63b	family with sequence similarity 63, member B
Taz	Tafazzin
Drosha	ribonuclease III, nuclear
Rps3a	ribosomal protein S3A
Myl9	myosin, light polypeptide 9, regulatory
Rpl6	ribosomal protein L6
Narf	nuclear prelamin A recognition factor
Arpc1b	actin related protein 2/3 complex, subunit 1B
Psmb4	proteasome (prosome, macropain) subunit, beta type 4
Polr1d	polymerase (RNA) I polypeptide D
Rpl10	ribosomal protein 10

Fth1	Ferritin Heavy Chain 1
Anxa5	Annexin A5
Plscr3	Phospholipid Scramblase 3
Wisp2	WNT1 Inducible Signaling Pathway Protein 2
Cops7a	COP9 Signalosome Subunit 7A
RanGAP1	Ran GTPase Activating Protein 1
TDG	Thymine DNA Glycosylase
PTP1B	Protein-tyrosine phosphatase-1B
GluR6	Glutamate Receptor 6

Chapter 1.

General Introduction

1-1. Post-translational modification in living cells.

In living cells, proteins are synthesized based on the information of DNAs (a process named translation). After the translation, proteins not only function by themselves, but also are modified by some chemical or protein modifiers to alter their functions diversely. Many kinds of post-translational protein modifications, such as phosphorylation, acetylation, methylation, and ubiquitination, have been reported, and specific residues including serine, threonine, and arginine in substrate proteins are targeted for the covalent modification[1]. The post-translational protein modification is an essential biological phenomenon to sustain cell viability.

1-1-1. Post-translational modification "SUMOylation".

In 1996, modification of a Small Ubiquitin-related Modifier (SUMO) protein, named SUMOylation, was newly discovered[2]. SUMO proteins are ubiquitously expressed by all eukaryotes and have several isoforms, the number of which is distinct between species[3–9]. Lower eukaryotes like yeast and *C. elegans* have a single SUMO gene, whereas higher eukaryotes like plants and vertebrates have multiple SUMO genes. Mammalians including human and mouse are known to express at least 3 SUMO genes: SUMO-1, SUMO-2, and SUMO-3. SUMO-2 and SUMO-3 are grouped together as SUMO-2/3 because the amino acid sequences of SUMO-2 and SUMO-3 are 97% homologous to each other[3,5,10]. On the other hand, the homology of the amino acid sequences between SUMO-1 and SUMO-2/3 is only 50%[3,5,10]. The SUMO family members covalently modify substrate proteins, forming an isopeptide bond between C-terminal glycine residue of the SUMO and lysine residue in the substrate (Fig. 1-1)[3,5,10]. In most cases, SUMO binds to the lysine residue in the motif "Ψ-K-X-E/D" (Ψ: a hydrophobic amino acid, K: the SUMO-modified lysine residue, X: any amino acid, E: a glutamic acid, D: an aspartic acid), which is called a SUMO consensus recognition site[3,5,10,11]. Another case has been reported; some proteins are SUMOylated at the sites different from

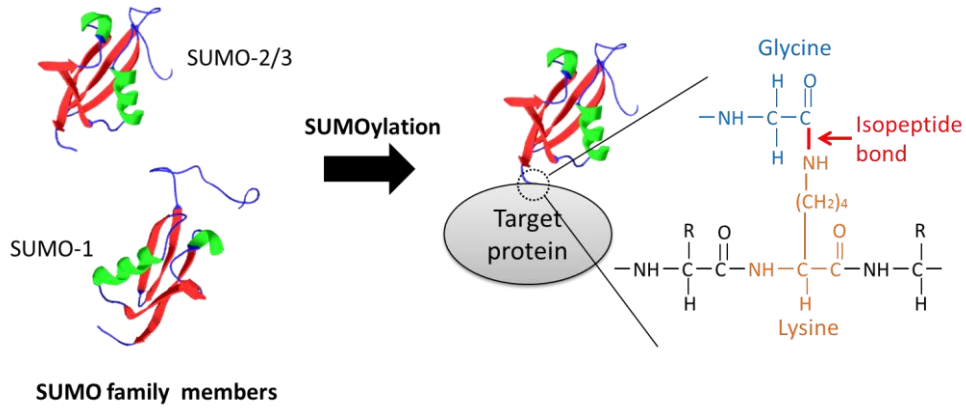

Fig. 1-1. Schematic of SUMOylation.

An isopeptide bond is formed between a SUMO and a target protein in SUMOylation reaction.

the SUMO consensus recognition sites[3]: PML is SUMOylated at "AKCP" sequences[12] and Smad4 at "VKYC" sequences[13]. SUMO-2/3 has the consensus recognition site in its amino terminal region and can be self-SUMOylated, forming polySUMO chains (Fig. 1-2), whereas SUMO-1 unlikely to form polySUMO chains[3,10]. The SUMO family members show different substrate selectivity: some substrate proteins are equally modified by all the SUMO family members; other substrate proteins are preferentially modified by one of the members[14,15]. The famous example is RanGAP1, a protein firstly discovered to be SUMOylated[2], and preferentially modified by SUMO-1[16]. The mechanism of the substrate selectivity has not been clarified yet, but it is reportedly suggested that the SUMO-ligases required for the SUMOylation process are partially related to select the substrates[10,14,15]. The fact that SUMOylation occurs without E3s in *in vitro* SUMOylation assay[17,18] also suggests that the variety of E3s is related to SUMO-paralogue specificity. The effectivity of SUMOylation is sometimes affected by mammalian cell types[19,20]. For instance, Ets-1, a transcription factor protein, is SUMOylated by all the SUMO family members in RK13 (Rabbit kidney) cells, but is more efficiently SUMOylated by SUMO-2 than SUMO-1 and SUMO-3 in COS-7 (Monkey kidney) cells and HEK293T (Human embryonic kidney) cells[19]. Stress stimulation

Human SUMO isoforms	N-terminal amino acid sequences (Amino acids from 1 to 20 are shown)
SUMO-1	MSDQEAKPSTEDLGDKKEGE
SUMO-2	MADEKPKEGVKTENNDHINL
SUMO-3	MSEEKPKEGVKTENDHINLK

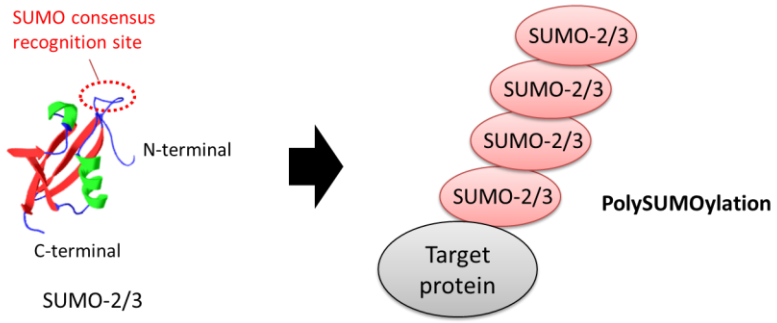

Fig. 1-2. Schematic of polySUMOylation.

N-terminal amino acid sequences from 1 to 20 of human SUMO isoforms are shown in the upper table. Red characters show the SUMO consensus recognition sites. Location of the SUMO consensus recognition site on the SUMO-2/3 structure and a simplified diagram of polySUMOylated form are shown in the bottom.

such as UV exposure, heat shock, and osmotic change also modulates SUMOylation in both promotion and suppression manner, especially the modification by SUMO-2/3[21].

1-1-2. Reaction mechanism of SUMOylation.

SUMOylation is a reversible process, which is mediated by three kinds of SUMOylation enzymes and several kinds of deSUMOylation enzymes (Fig. 1-3)[3,4,22]. The glycine residue, which binds to the lysine residue in a target protein, is not exposed in a newly-synthesized SUMO. SUMO/sentrin-specific proteolytic enzymes (SENP) remove some carboxy-terminal amino acids of a newly-synthesized SUMO, which result in exposure of the glycine residue. Next, a SUMO-activating enzyme (E1) carries out an ATP-dependent activation of a SUMO. A high-energy thiolester bond is formed between a cysteine in the E1 and the C-terminus of a SUMO. Then, SUMO is transferred and conjugated to a cysteine in the SUMO-conjugating enzyme (E2). Subsequently,

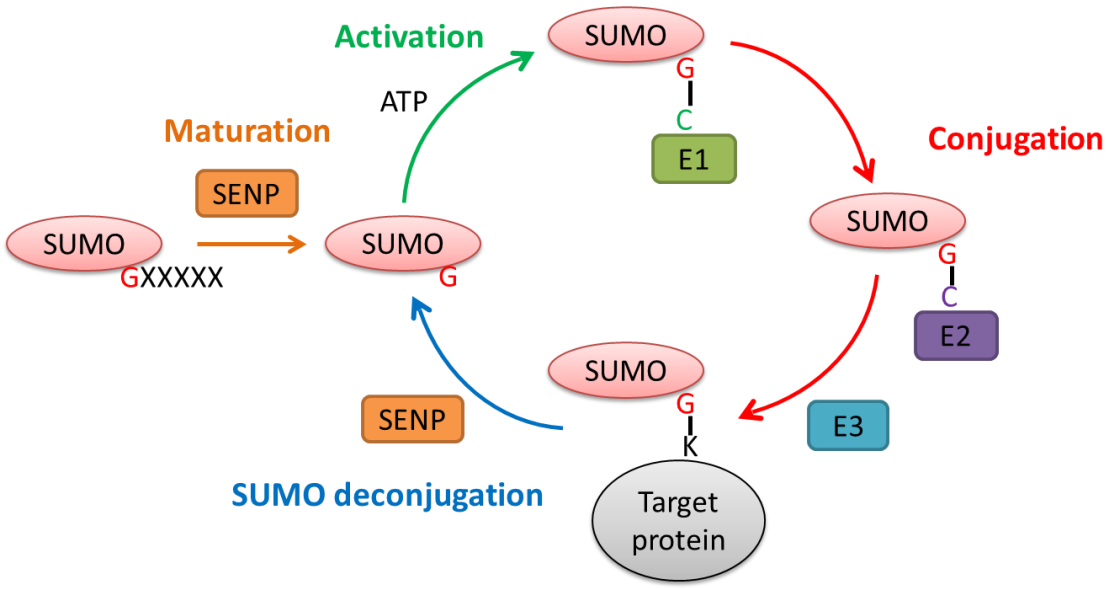

Fig. 1-3. Schematic of the SUMOylation mechanism.

A reversible SUMOylation process is mediated by a SUMO/sentrin-specific protease (SENP) and three kinds of SUMOylation enzymes, a SUMO-activating enzyme (E1), a SUMO-conjugating enzyme (E2), and a SUMO-protein ligase (E3). “G”, “C”, and “K” indicate Glycine, Cysteine, and Lysine, respectively. “XXXXX” represents a C-terminal amino acid sequence of SUMO.

SUMO is transferred to a substrate protein and an isopeptide bond is formed between the SUMO and the substrate protein, the process of which is assisted with a SUMO-protein ligase (E3). The SENPs also cleave the isopeptide bond, and the released SUMO is again used for SUMOylation.

Most organisms contain a single E1 enzyme, which is a heterodimer of Aos1 and Uba2, and a single E2 enzyme known as Ubc9 protein, which directly binds to the SUMO consensus recognition site “ΨKXE”. In contrast, there are several kinds of E3 enzymes different between organisms[3]. For example, yeasts have two E3s, Siz1 and Siz2. Mammalians have 3 types of E3s: protein inhibitor of activated STAT (PIAS) family; nuclear pore protein RanBP2/Nup358; polycomb group protein Pc2. Also, various SUMO proteases different between organisms have been identified[23]. In yeast, Ulp1 and Ulp2 are known as SUMO proteases. Mammalians have several types of SUMO proteases: sentrin-specific protease (SENP) class, SENP1-3 and SENP5-7; desumoylating isopeptidase 1

(DESI1) and DECI2; USPL1. The SENP family members have been reported to have different SUMO-paralogue preferences and activities[23].

1-1-3. SUMO-related cellular functions.

SUMOylation is considered to alter the functions of substrate proteins by the following three ways at the molecular level[5]. The first way is that SUMOylation creates a new binding site on the surface of the substrate protein (Fig. 1-4A). The second way is that SUMOylation induces an intramolecular conformational change of the substrate protein via SUMO-interacting motifs (SIMs), such as [V/I]-X-[V/I]-[V/I] ("V": valine, "I": isoleucine, "X": one of any amino acids), with which SUMO interacts non-covalently (Fig. 1-4B)[24]. The third way is that SUMOylation masks the surface on the substrate proteins, resulting in prevention of an interaction with other proteins (Fig. 1-4C). The

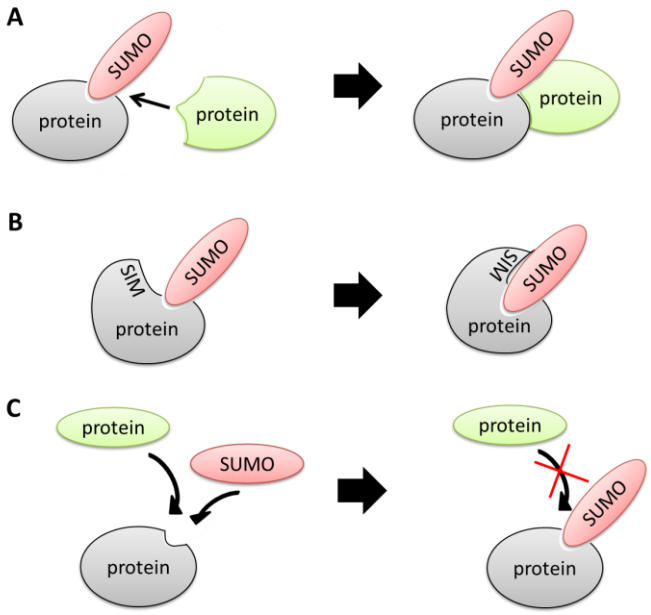

Fig. 1-4. Schematic of modulation of a substrate protein by SUMOylation at a molecular level.

A. Creation of a new binding surface on a substrate protein by SUMO binding. B. Induction of intra-molecular conformational change through non-covalent interaction between a SUMO-interacting motif (SIM) on a substrate protein and SUMO. C. Blocking of an interaction of a substrate protein with another protein by SUMO attachment to the binding site.

change of the surface condition on substrate proteins by SUMOylation is considered to influence the interaction with other proteins, which is fundamental to induce cellular functions.

Proteins with various cellular localizations including cytosol and nucleus are targets for SUMOylation. The substrate proteins have been reported to be implicated in various cellular phenomena such as DNA repair, transcription, circadian, and signal transduction, which are indispensable for cell sustainability (Table. 1-1)[3–5,25–33]. Many proteins related to diseases such as cancer, cardiomyopathy, Alzheimer's diseases, and Parkinson's diseases have been also reported to be SUMO-targets, suggesting that SUMOylation is participated in the diseases pathogenesis[34]. Therefore, SUMOylation plays an important role in regulating protein functions variously throughout a whole cell.

Table. 1-1. Examples of SUMO-related phenomena

Protein	Protein localization	SUMO-related phenomena	Reference
RanGAP1	Cytosol	Translocation from cytosol to periphery of a nucleus.	35
TDG	Nucleus	DNA repair	32
Elk-1	Nucleus	Transcription	31
Bmal1	Nucleus	Circadian	28
GluR6	Plasma membrane	Endocytosis	33
PTP1B	Endoplasmic reticulum	Receptor tyrosine kinase signaling	36
Lmna	Nuclear periphery	Abnormal SUMOylation is related to familial dilated cardiomyopathy.	37

1-2. The conventional screening methods for mammalian or yeast SUMOylated proteins.

Mammalian and yeast SUMOylated proteins have been screened by various methods as follows. These methods have several advantageous points for the detection of SUMOylation. However, there are still some difficulties in detecting mammalian SUMOylated proteins.

1-2-1. Immunoprecipitation-based mass spectrometric screening method for mammalian SUMOylated proteins using mammalian cells.

The discovery of novel SUMOylated proteins is significant for the deeper understanding of the SUMOylation roles in diverse biological events. However, detection of SUMOylated proteins is considerably difficult, because very small portion of target proteins, generally less than 1%, are SUMOylated at steady state and SUMO proteins are easily and rapidly removed from substrates by deSUMOylation enzymes upon cell lysis[3]. Conventionally, immunoprecipitation method followed by mass spectrometric analysis (IP-MS) is used to screen mammalian SUMOylated proteins[38,39]. In the method, SUMO fused with an epitope tag is exogenously expressed in mammalian cells. Then, the mammalian cells are lysed, and the endogenous proteins, which are modified by the exogenously-expressed SUMO proteins, are immunoprecipitated from the cell lysates using antibodies specific to the tag attached to SUMO. The collected proteins are digested by trypsin, and the resultant peptides are subjected to MS analysis to identify SUMOylated proteins (Fig. 1-5). In the IP procedure, the cell lysis is performed in a denaturing condition to inhibit deSUMOylation. However, complete inhibition of the SUMO proteases is difficult because strong denaturing condition disrupts the antibody activity. Thus, SUMO proteins can be removed from substrate proteins during the IP procedure. This possibility causes biased detection of SUMOylated proteins, which are highly SUMOylated and not rapidly deSUMOylated in cells. Thereby, there are some

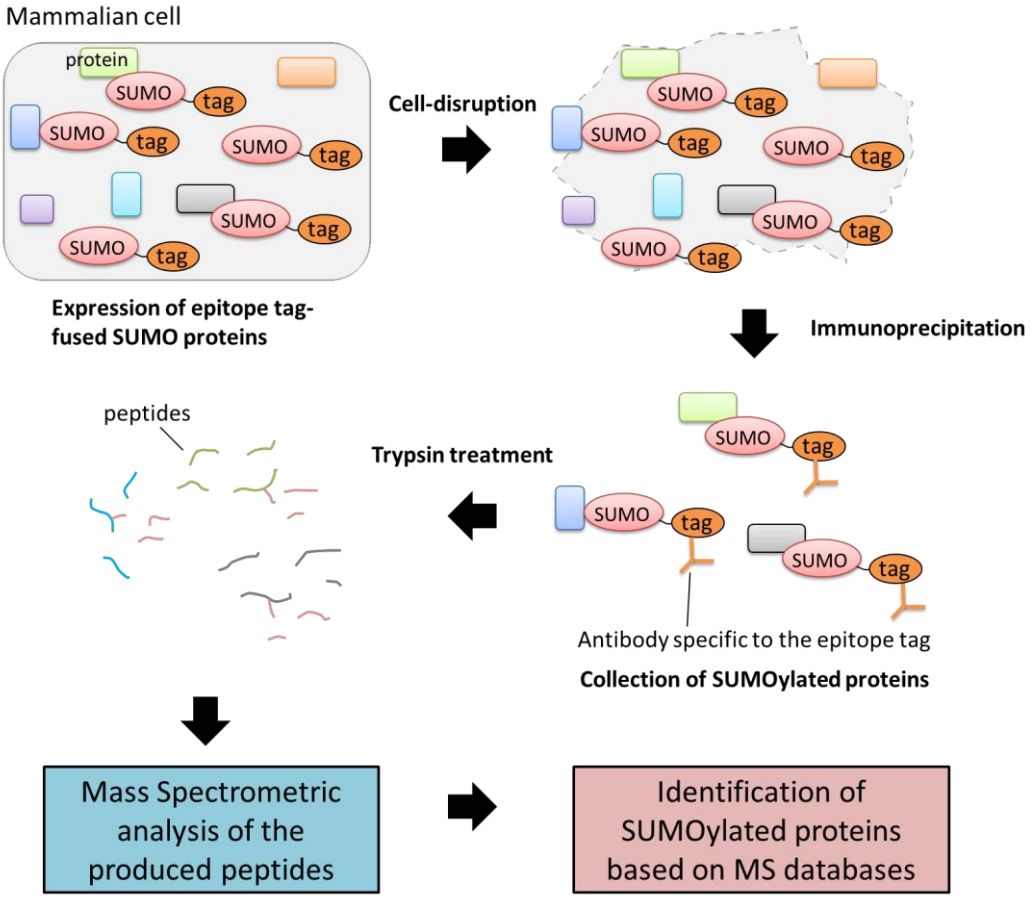

Fig. 1-5. Schematic of an immunoprecipitation-based mass spectrometric screening method for mammalian SUMOylated proteins.

Epitope-tagged SUMO proteins are exogenously expressed in mammalian cells. Substrate proteins modified by the SUMO are collected from cell lysate by immunoprecipitation with an antibody specific to the epitope tag. The immunoprecipitated proteins are subjected to trypsin treatment and the resultant peptides are analyzed by mass spectrometry to identify the SUMOylated proteins.

limitations in the scope of detectable SUMOylation candidates in the IP-MS screening method.

1-2-2. Yeast two-hybrid screening method for yeast SUMOylated proteins using yeast cells.

Yeast two-hybrid screening was devised to screen yeast SUMOylated proteins (Fig. 1-6)[40]. In the basic principle of the yeast two-hybrid method, a transcriptional factor is divided into two fragments, a DNA-binding domain and a transcriptional activation domain. Each fragment is genetically fused to a yeast SUMO and a protein that is coded by a yeast cDNA library (named a library protein), respectively, and both proteins are co-expressed in individual yeast cells. When the two fragments come close together upon SUMOylation of the library protein, the transcriptional activity of the divided transcription factor recovers. This results in transcription of the downstream reporter gene. The expressed reporter protein reacts with the materials included in the culture medium, by which the yeasts get colored blue due to the production of blue pigment. Hence, the yeasts, in which library proteins are SUMOylated, can be selected by the expression of the reporter gene, and the library proteins can be identified by analyzing the DNA sequences that code the library proteins. Unlike the immunoprecipitation method, the yeast-two hybrid method does not require a cell-disruption process. However, the method requires translocation of SUMO and a library protein into nucleus. Thus, proteins that are difficult to translocate into the nucleus are improper as targets for evaluating SUMOylation.

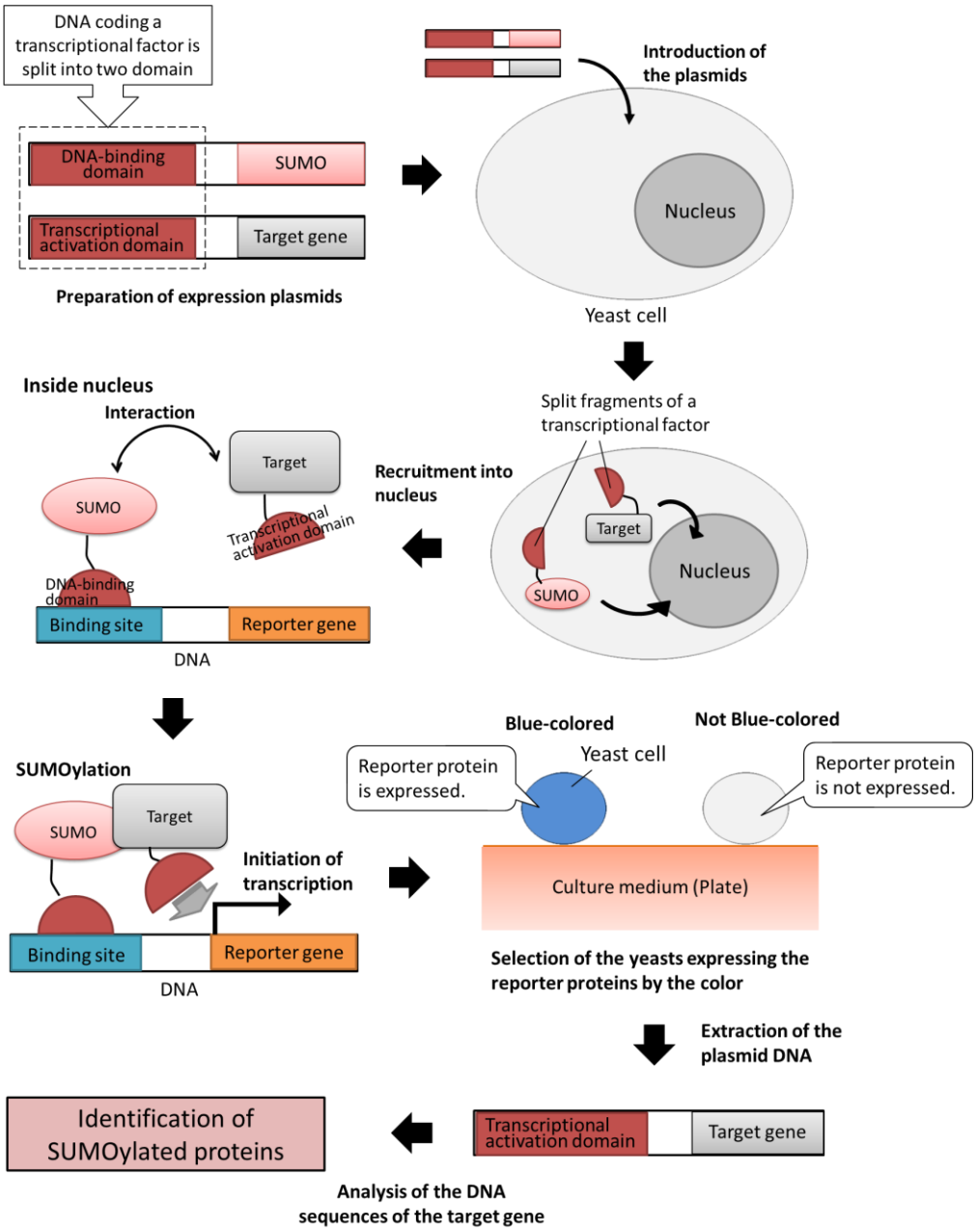

Fig. 1-6. Schematic of a yeast-two hybrid screening method for yeast SUMOylated proteins.
A fragment including a DNA-binding domain of a transcriptional factor is fused with yeast SUMO. Another fragment including a transcriptional activation domain of the transcriptional factor is fused to a target protein. In nucleus, the fragment including the DNA-binding domain binds to a specific region on DNA. When the target protein is modified by the SUMO, the two fragments close together. This results in recovery of the transcriptional activity and the downstream reporter gene is started to be transcribed. Only the yeasts in which the reporter proteins are expressed get colored blue. DNAs are extracted from the yeasts and the DNA sequences coding the library proteins are analyzed to identify the SUMOylated proteins.

1-2-3. Fluorescence protein reconstitution-based screening method for yeast SUMOylated proteins using yeast cells.

Recently, the non-cell-disruptive screening method for yeast SUMOylated proteins that does not limit the subcellular location of SUMOylation has been devised[41] (Fig. 1-7). In the method, reconstitution of split fluorescence protein fragments is used. A yeast SUMO and a protein coded by yeast cDNA library are genetically fused to each split fluorescent fragment, which lose the ability of emitting fluorescence. Each protein is expressed in different types of yeast cells; "α" type and "a" type. By mating the "α" type yeast with the "a" type, a cell-fusion occurs, which results in coexpression of the SUMO and the library protein in one yeast cell. If the library protein is modified by the SUMO, the fused fluorescence protein fragments come close to each other, resulting in reconstitution of the split fluorescence protein fragments and recovery of fluorescence emission. Thus, SUMOylation can be detected as a fluorescence signal. The yeasts containing different library proteins are individually plated on each well of multi-well plates with the yeast containing the SUMO. SUMOylation of individual library proteins is assessed and SUMOylated proteins are identified. This method is advantageous in being detectable of yeast SUMOylated proteins without a cell-disruption process. In addition, both a SUMO and a library protein are not required to translocate to a certain subcellular compartment. A mammalian SUMO and a mammalian cDNA library can be applied to this method. However, there are several problems in the detection of mammalian SUMOylation in yeast cells. First, yeast SUMOylation system does not completely reflect mammalian SUMOylation system: yeast cells have single SUMO isoform[8,9], whereas the mammalian cells express at least three SUMO isoforms with different substrate specificity[14]; SUMOylation patterns of mammalians, the multi-cellular organisms, sometimes differ upon the cell types[19,20], and the cell-type-specific SUMOylation cannot be assessed using yeasts, the single-cellular organisms. Therefore, it is desired to detect mammalian SUMOylation in living

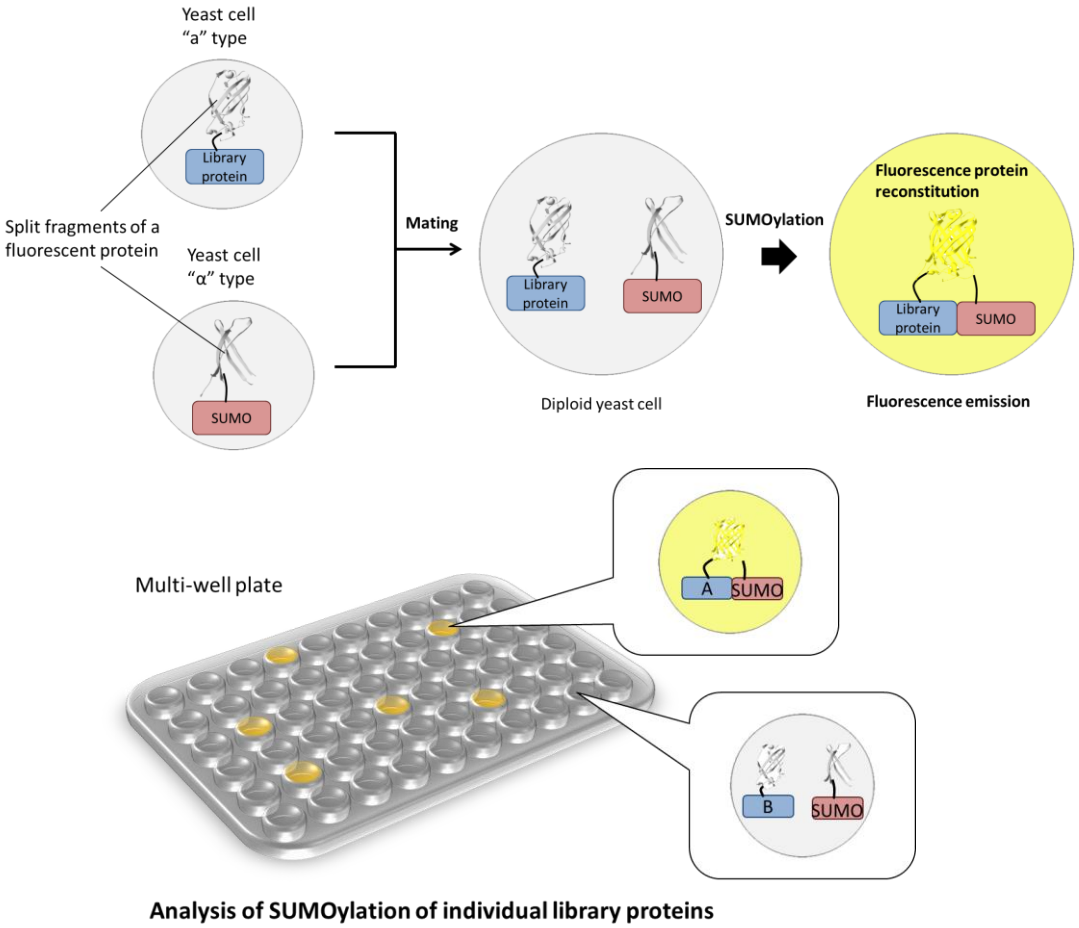

Fig. 1-7. Schematic of a fluorescence protein reconstitution-based screening method for yeast SUMOylated proteins.

A fluorescence protein is divided into non-fluorescent fragments. Each fragment is genetically fused to a yeast SUMO and a target protein encoded by a yeast cDNA library. The two proteins are expressed in different yeast cells, an “α” type yeast cell and an “a” type yeast cell. By mating the two different yeast cells with each other, a diploid yeast cell containing the both proteins is generated. When the library protein is modified by the SUMO, the fused fluorescence protein fragments get close to each other, which results in fluorescence protein reconstitution and recovery of the fluorescence emission. SUMOylation of individual library proteins is assessed using multi-well plates.

mammalian cells.

In this method, the yeast containing the SUMO is mated with the yeast containing the library proteins to generate the diploid yeast cell containing both the SUMO and the library protein. Only

the combination of “α” type yeast and “a” type yeast causes the cell-fusion. This means that the yeast cell containing the SUMO does not mate with the same type of a yeast cell containing the SUMO, as is the same case with the yeast cell containing the library protein. Hence, the mating of the two yeast cell types is beneficial in efficiently introducing both proteins in one yeast cell. In this system, mammalian cells cannot be simply substituted for the yeast cells because mammalian cell-fusion is not easy to handle compared to the yeast cell-fusion and control of selectively fusing a mammalian cell containing a SUMO with a mammalian cell containing a library protein is difficult. Thus, another approach is required for the detection of mammalian SUMOylation.

1-3. Techniques required for the detection of SUMOylation in living mammalian cells.

Observation of proteins under a live-cell condition is significant for gaining insights into biological mechanisms. For detection of target proteins in living cells, fluorescence proteins have been widely used. By generically fusing fluorescence proteins to target proteins and introducing the plasmid DNA coding the fusion protein, the target proteins can be visualized as fluorescence signals in living mammalian cells by a fluorescence microscopy. For detection of protein-protein interaction, reconstitution of split fluorescence protein fragments has been used[42–44]. By fusing split fluorescence protein fragments to target two proteins, the interaction between the two target proteins brings the fused split fluorescence protein fragments together, resulting fluorescence protein reconstitution and recovery of fluorescence emission. The reconstitution reaction is irreversible. This is beneficial in detecting transient interaction between two proteins. Once reconstitution of split fluorescence protein fragments occurs, the fluorescence signal can be maintained after the two proteins are detached. Hence, the fluorescence protein reconstitution is suitable for detecting SUMOylation, the reversible and low-level protein modification. Using the reconstitution technique, our laboratory has developed screening methods for the proteins localized in mitochondria or endoplasmic reticulum[45,46]. In the screening, a fluorescence-activated cell sorter (FACS) is used. This machine can automatically sort target fluorescent mammalian cells in a rapid and high throughput manner under a live-cell condition. Hence, FACS is suitable for collecting the mammalian cells that contain reconstituted fluorescence proteins induced by SUMOylation.

1-3-1. Basic characteristic of a fluorescence protein.

Fluorescent proteins such as Green Fluorescent Protein (GFP) emit fluorescence from the chromophore, which is surrounded by β-barrel fold[47,48]. The chemical structure of the chromophore is different between fluorescent protein species (an example of GFP chromophore is shown in Fig. 1-8A). The chromophore is auto-catalytically formed through multiple reaction steps, such as cyclization and oxidation (Fig. 1-8B). For the beginning of the chromophore maturation, the correct protein folding of the β-barrel structure is indispensable.

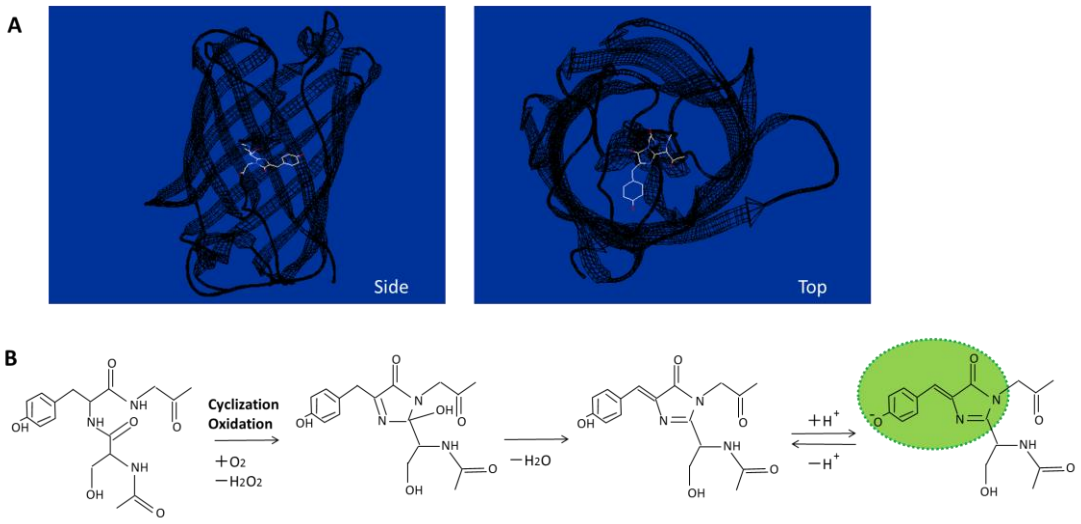

Fig. 1-8. Characteristics of a GFP chromophore.

A. The GFP chromophore is inside the β-barrel fold. The GFP structures are shown as black. B. Schematic of GFP chromophore maturation. A GFP chromophore matures through several steps, including cyclization and oxidation. The green circle indicates the chemical structure with an expanded system of π-conjugated electrons, where visible light is absorbed and fluorescence is emitted.

1-3-2. Basic principle of fluorescence protein reconstitution.

When a fluorescent protein is dissected into two fragments, it loses the ability of emitting fluorescence. However, when the two fragments are brought into proximity, reconstitution of the fluorescent protein occurs, resulting in recovery of the fluorescence. By fusing two proteins of interest to the split fluorescent protein fragments, respectively, interaction between the two proteins brings about the proximity of the attached fragments, which induces reconstitution of the fluorescence protein and recovery of fluorescence emission (Fig. 1-9). Hence, the interaction between the two proteins can be detected as fluorescence in living cells. This reconstitution of fluorescent protein is irreversible, thereby being useful to detect transient or weak protein-protein interactions.

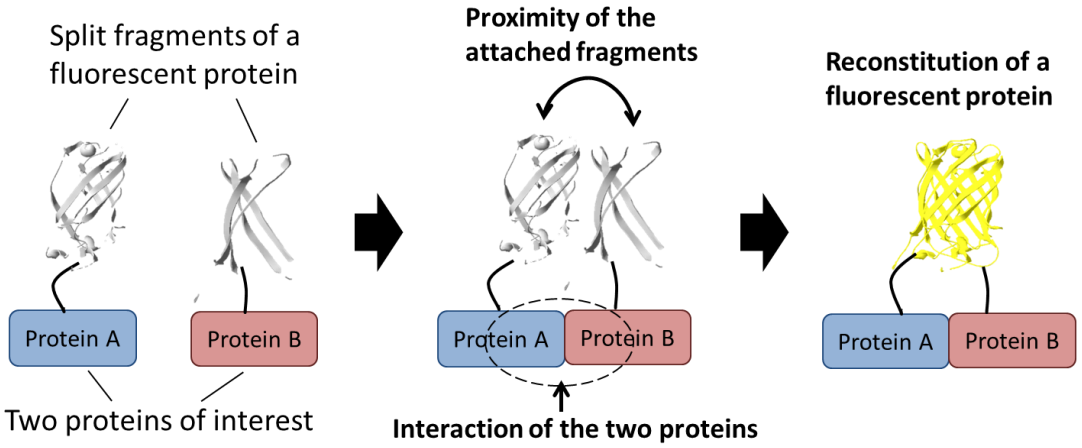

Fig. 1-9. Principle of reconstitution of split fluorescent protein fragments.
A fluorescent protein is divided into two non-fluorescent fragments, which are fused to two proteins respectively. When the two proteins interact with each other, the attached fragments are brought into proximity, resulting in reconstitution of the fluorescent protein and recovery of the fluorescence emitting.

For the appropriate reconstitution of split fluorescence protein fragments, the split position of amino acid in fluorescence proteins is important. Many studies have been conducted to reveal the proper split positions in various fluorescence proteins, such as a green fluorescence protein (GFP) and an enhanced yellow fluorescence protein (EYFP). The reported split positions in several fluorescence proteins are summarized in Table 1-2[49–51].

Table 1-2. Split positions in various fluorescence proteins for the reconstitution.

Fluorescence protein	Excitation (nm)	Emission (nm)	Split position (amino acid)
GFP	485	500	155-156 157-158
EGFP	484	507	157-158
BFP	380	440	172-173
ECFP	452	478	154-155 172-173
EYFP	515	527	154-155 172-173
Venus	515	528	144-145 154-155 158-159 172-173
mRFP	549	570	154-155 168-169
mCherry	587	610	159-160
Citrine	516	529	154-155 172-173
mKate	588	625~635	151-152 165-166
mLumin	587	621	151-152 165-167

1-3-3. Basic principle of Fluorescence-activated cell sorter (FACS).

A fluorescence-activated cell sorter (FACS) is an equipment to detect fluorescent intensities of a cell population, and the collected data are displayed as a histogram (an example is shown in Fig. 1-10A). In addition, it enables to sort target fluorescent cells. The basic principle of the cell-sorting is shown in Fig. 1-10B. A sheath solution runs in a FACS. A sample solution including cells flows in the sheath solution. Hydrodynamic effect by the pressure of the sheath solution makes the cells flow one by one. Then, vibration makes the sample solution form drops, which individually have a single cell. Laser comes from a horizontal side and the fluorescent intensities of individual cells are detected by a detector. The drops including target fluorescent cells are electrically charged. The charged drops are electrically attracted by a deflection plate, by which the target cells are separated.

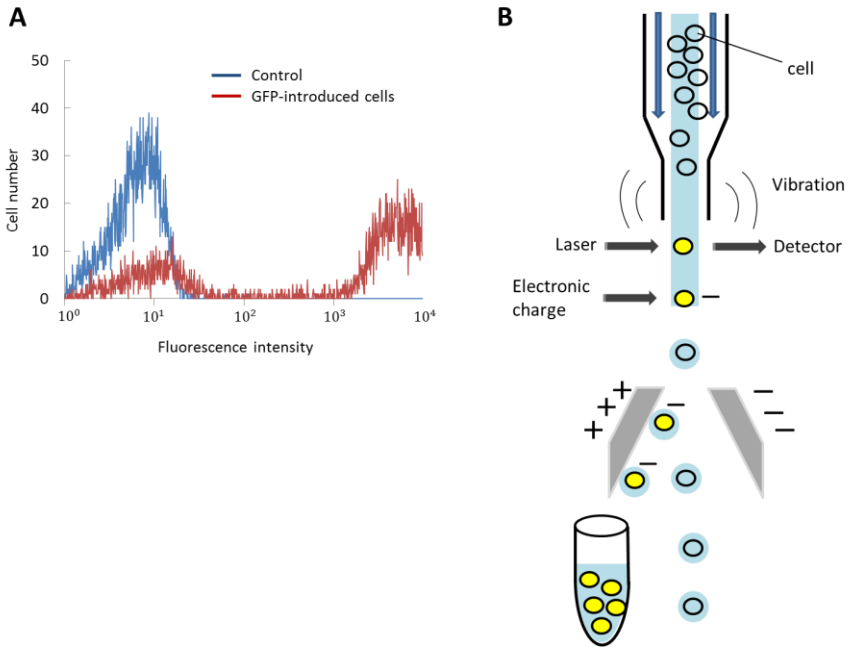

Fig. 1-10. Characteristics of a fluorescence-activated cell sorter (FACS).

A. Detection of fluorescent intensities of a specific cell population by FACS. As examples, fluorescent intensities of 5,000 control cells (Blue) and 5,000 GFP-introduced cells (Red) are shown in the histograms. B. Schematic of principle of cell sorting by FACS. Cells flow one by one. The fluorescent intensities of individual cells are detected. Based on the detected fluorescent intensities, target fluorescent cells are electrically charged and collected by a deflection plate with the opposite electric charges.

1-4. Purpose of the present study.

In the conventional methods, there are still some limitations in detection of mammalian SUMOylated proteins. In the immunoprecipitation-based method, a cell-disruption process potentially causes deSUMOylation by SUMO proteases and SUMOylated proteins that are highly and frequently modified are preferentially immunoprecipitated from the cell lysates. On the other hand, in the yeast-two-hybrid-based method, location of SUMOylation is limited in a nucleus and yeast cells cannot completely reflect SUMOylation condition in mammalian cells. Therefore, to overcome these difficulties, a new approach to identify mammalian SUMOylated proteins is required. In this study, I aimed to develop a non-destructive screening method using living mammalian cells without limitation in subcellular protein locations for assessing SUMOylation. To detect SUMOylation under a live-cell condition, the reconstitution of split fluorescent protein fragments[52,43,42,46] was used. This reconstitution technique is useful to detect protein-protein interaction without cell-disruption. The interaction could be detected by fluorescence emission. An irreversible process of the reconstitution of split fluorescent protein fragments is adequate for detecting the reversible SUMOylation in that SUMOylation signals could be maintained even though deSUMOylation occurs later. By combining the reconstitution technique with fluorescence-activated cell sorter (FACS), which automatically collects cells emitting fluorescence, the cells that harbor library proteins modified by SUMOs can be separated in a high-throughput manner. Thus, the reconstitution of the fluorescent protein fragments meets the requirements in the detection of infrequent SUMOylation in living mammalian cells.

I herein describe a development of a non-destructive screening system for SUMOylated proteins, which is based on the reconstitution of split fluorescent protein fragments combined with a fluorescence-activated cell sorter (FACS), using living mammalian cells and also describe the results of screening by the present method.

Reference

1. Yang, X.-J. Multisite protein modification and intramolecular signaling. *Oncogene* **24,** 1653–1662 (2005).
2. Matunis, M. J., Coutavas, E. & Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **135,** 1457–1470 (1996).
3. Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–82 (2004).
4. Melchior, F. SUMO--nonclassical ubiquitin. *Annu. Rev. Cell Dev. Biol.* **16,** 591–626 (2000).
5. Geiss-friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* **8,** 947–956 (2007).
6. Lapenta, V. *et al.* SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to chromosome 21qter and defines a novel gene family. *Genomics* **40,** 362–366 (1997).
7. Chen, A., Mannen, H. & Li, S. S. Characterization of mouse ubiquitin-like SMT3A and SMT3B cDNAs and gene/pseudogenes. *Biochem. Mol. Biol. Int.* **46,** 1161–1174 (1998).
8. Takahashi, Y., Toh-E, A. & Kikuchi, Y. Comparative analysis of yeast PIAS-type SUMO ligases in vivo and in vitro. *J. Biochem.* **133,** 415–422 (2003).
9. Bylebyl, G. R., Belichenko, I. & Johnson, E. S. The SUMO Isopeptidase Ulp2 Prevents Accumulation of SUMO Chains in Yeast. *J. Biol. Chem.* **278,** 44113–44120 (2003).
10. Gareau, J. R. & Lima, C. D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat. Rev. Mol. Cell Biol.* **11,** 861–871 (2010).
11. Rodriguez, M. S., Dargemont, C. & Hay, R. T. SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J. Biol. Chem.* **276,** 12654–12659 (2001).
12. Kamitani, T. *et al.* Identification of three major sentrinization sites in PML. *J. Biol. Chem.* **273,** 26675–26682 (1998).

13. Lee, P. S. W., Chang, C., Liu, D. & Derynck, R. Sumoylation of Smad4, the common Smad mediator of transforming growth factor-beta family signaling. *J. Biol. Chem.* **278,** 27853–27863 (2003).

14. Vertegaal, A. C. O. *et al.* Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol. Cell. Proteomics* **5,** 2298–2310 (2006).

15. Tatham, M. H. *et al.* Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. *Nat. Struct. Mol. Biol.* **12,** 67–74 (2005).

16. Zhu, S. *et al.* Protection from isopeptidase-mediated deconjugation regulates paralog-selective sumoylation of RanGAP1. *Mol. Cell* **33,** 570–580 (2009).

17. Okuma, T., Honda, R., Ichikawa, G., Tsumagari, N. & Yasuda, H. In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. *Biochem. Biophys. Res. Commun.* **254,** 693–698 (1999).

18. Desterro, J. M., Rodriguez, M. S., Kemp, G. D. & Hay, R. T. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *J. Biol. Chem.* **274,** 10618–10624 (1999).

19. Ji, Z. *et al.* Regulation of the Ets-1 transcription factor by sumoylation and ubiquitinylation. *Oncogene* **26,** 395–406 (2007).

20. Degerny, C. *et al.* SUMO modification of the Ets-related transcription factor ERM inhibits its transcriptional activity. *J. Biol. Chem.* **280,** 24330–24338 (2005).

21. Tempé, D., Piechaczyk, M. & Bossis, G. SUMO under stress. *Biochem. Soc. Trans.* **36,** 874–878 (2008).

22. Capili, A. D. & Lima, C. D. Taking it step by step: mechanistic insights from structural studies of ubiquitin/ubiquitin-like protein modification pathways. *Curr. Opin. Struct. Biol.* **17,** 726–735 (2007).

23. Hickey, C. M., Wilson, N. R. & Hochstrasser, M. Function and regulation of SUMO proteases.

Nat. Rev. Mol. Cell Biol. **13,** 755–766 (2012).

24. Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **281,** 16117–16127 (2006).

25. Ouyang, J. & Gill, G. SUMO engages multiple corepressors to regulate chromatin structure and transcription. *Epigenetics* **4,** 440–444 (2009).

26. Bergink, S. & Jentsch, S. Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* **458,** 461–467 (2009).

27. Pelisch, F. *et al.* Dynamic SUMO modification regulates mitotic chromosome assembly and cell cycle progression in Caenorhabditis elegans. *Nat. Commun.* **5,** 5485 (2014).

28. Lee, J. *et al.* Dual modification of BMAL1 by SUMO2/3 and ubiquitin promotes circadian activation of the CLOCK/BMAL1 complex. *Mol. Cell. Biol.* **28,** 6056–6065 (2008).

29. Girdwood, D. *et al.* P300 transcriptional repression is mediated by SUMO modification. *Mol. Cell* **11,** 1043–1054 (2003).

30. Myatt, S. S. *et al.* SUMOylation inhibits FOXM1 activity and delays mitotic transition. *Oncogene* **33,** 4316–4329 (2014).

31. Yang, S.-H., Jaffray, E., Hay, R. T. & Sharrocks, A. D. Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. *Mol. Cell* **12,** 63–74 (2003).

32. Hardeland, U., Steinacher, R., Jiricny, J. & Schär, P. Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J.* **21,** 1456–1464 (2002).

33. Martin, S., Nishimune, A., Mellor, J. R. & Henley, J. M. SUMOylation regulates kainate-receptor-mediated synaptic transmission. *Nature* **447,** 321–325 (2007).

34. Sarge, K. D. & Park-Sarge, O.-K. Sumoylation and human disease pathogenesis. *Trends Biochem. Sci.* **34,** 200–205 (2009).

35. Matunis, M. J., Wu, J. & Blobel, G. SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *J. Cell Biol.* **140,** 499–509 (1998).

36. Dadke, S. *et al.* Regulation of protein tyrosine phosphatase 1B by sumoylation. *Nat. Cell Biol.* **9,** 80–85 (2007).

37. Zhang, Y.-Q. & Sarge, K. D. Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies. *J. Cell Biol.* **182,** 35–39 (2008).

38. Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K. & White, M. A. Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J. Biol. Chem.* **279,** 20999–21002 (2004).

39. Tirard, M. *et al.* In vivo localization and identification of SUMOylated proteins in the brain of His6-HA-SUMO1 knock-in mice. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21122–21127 (2012).

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

41. Sung, M.-K. *et al.* Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. *Genome Res.* **23,** 736–746 (2013).

42. Hu, C.-D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9,** 789–798 (2002).

43. Kerppola, T. K. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat. Protoc.* **1,** 1278–1286 (2006).

44. Hu, C.-D. & Kerppola, T. K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* **21,** 539–545

(2003).

45. Ozawa, T., Nishitani, K., Sako, Y. & Umezawa, Y. A high-throughput screening of genes that encode proteins transported into the endoplasmic reticulum in mammalian cells. *Nucleic Acids Res.* **33,** e34 (2005).

46. Ozawa, T., Sako, Y., Sato, M., Kitamura, T. & Umezawa, Y. A genetic approach to identifying mitochondrial proteins. *Nat. Biotechnol.* **21,** 287–293 (2003).

47. Stepanenko, O. V, Stepanenko, O. V, Kuznetsova, I. M., Verkhusha, V. V & Turoverov, K. K. Beta-barrel scaffold of fluorescent proteins: folding, stability and role in chromophore formation. *Int. Rev. Cell Mol. Biol.* **302,** 221–78 (2013).

48. Kim, S. B. & Ozawa, T. Creating bioluminescent indicators to visualise biological events in living cells and animals. *Supramol. Chem.* **22,** 440–449 (2010).

49. Wehr, M. C. & Rossner, M. J. Split protein biosensor assays in molecular pharmacological studies. *Drug Discov. Today* **21,** 415–429 (2016).

50. Shyu, Y. J. & Hu, C.-D. Fluorescence complementation: an emerging tool for biological research. *Trends Biotechnol.* **26,** 622–630 (2008).

51. Herrera, F., Gonçalves, S. & Outeiro, T. F. Imaging protein oligomerization in neurodegeneration using bimolecular fluorescence complementation. *Methods Enzymol.* **506,** 157–174 (2012).

52. Isogai, M. *et al.* Structure and characteristics of reassembled fluorescent protein, a new insight into the reassembly mechanisms. *Bioorganic Med. Chem. Lett.* **21,** 3021–3024 (2011).

Chapter 2.

Development of a screening method for mammalian SUMOylated proteins

2-1. Introduction

SUMOylation plays important and versatile roles in living cells. The identification of novel SUMOylated proteins is prospective for unraveling new biological mechanisms regulated by SUMOylation. Conventionally, an immunoprecipitation-based Mass Spectrometric analytical method (IP-MS method) and a yeast-two hybrid method have been used to screen SUMOylated proteins. These methods have several limitations in detection of mammalian SUMOylated proteins. In the IP-MS method, a cell lysis process potentially causes deSUMOylation by SUMO proteases and detectable SUMOylated proteins would be biased because highly and frequently SUMOylated proteins are preferentially immunoprecipitated from the cell lysates. In the yeast-two hybrid method, translocation of both a SUMO and a target protein into a nucleus is required and yeast cells can not reflect mammalian intracellular condition and cell-type specific SUMOylation. In the present research, therefore, a new screening method was developed, where living mammalian cells were used during detection of SUMOylation and subcellular protein locations were not limited. This method was based on reconstitution of split fluorescent protein fragments and FACS sorting. The details of the process of the present method and the obtained results are described in this chapter.

2-2. Materials and Methods

2-2-1. Construction of expression plasmids

Library DNA series was prepared from pAP3neo cDNA Library Mouse 10T-half, 5.3M (RIKEN BRC DNA BANK). The cDNA was digested with *Eco*RI / *Not*I sites and then cloned into *Eco*RI / *Not*I sites of pMX vector. In the pMX vector, a DNA sequence corresponding to 159 to 240 amino acids of Venus, named VC, was cloned into *Bam*HI / *Eco*RI sites. Three different-length linkers, ggcggaggcgga, ggcggaggcggag, and ggcggaggcggagg, were amplified with the VC DNA by PCR using the following primers: 5’-TTTGGATCCGCTAGCGCCATGAAGAACGGCATCAAG-3’ and

5’-GGTGAATTC(/C/CC)TCCGCCTCCGCCCTTGTACAGCTCGTCC-3’. DNA of SUMO2 was amplified from the genome DNA of Hela cells using the following primers: 5’-GGCGGATCCATGGCCGACGAAAAGCCCAAG-3’ and 5’-CCCTCTAGACTCGAGCGGTAGACACCTCCCGTCTGCTG-3’. Then, the amplified SUMO2 DNA sequences were cloned into *Bam*HI / *Xho*I sites of pcDNA4/V5 His ver. B vector. In the pcDNA4/V5 His ver. B vector, DNA sequence corresponding to 1 to 158 amino acids of Venus, named VN, was cloned into *Hin*dIII / *Bam*HI sites using the following primers: 5’-AGTGCGGCCGCTTAAGCTTACCATGGTGAGCAAGGGCGAG-3’ and 5’-CATGGATCCGGCTCCGCCCTGCTTGTCGGCGGTGATATAG-3’ to generate the construct of VN-SUMO2. The DNA of RanGAP1 was amplified from RanGAP1 cDNA clone purchased from Kazusa DNA Res. Inst. using the following primers: 5’-TTTGGATCCGCCACCATGGCCTCGGAAGACATTGCCAAGCTGGCA-3’ and 5’-AAACTCGAGCTAGACCTTGTACAGCGTCTGCAGCAGACT-3’. Then, the amplified RanGAP1 gene was cloned into *Bam*HI / *Xho*I sites of pcDNA4/V5 His ver. B vector. In the pcDNA4/V5 His ver. B vector, DNA of VC was cloned into *Hin*dIII / *Bam*HI sites. Then, DNA of VC-RanGAP1 was extracted by digestion with *Hin*dIII / *Xho*I sites, and transferred into *Hin*dIII / *Xho*I sites of pMX vector. The DNAs that code partial RanGAP1 corresponding to 1 to 514, and 535 to 587 amino acids were respectively amplified using the following primers: 5’-TTTGGATCCGCCACCATGGCCTCGGAAGACATTGCCAAGCTGGCA-3’ and 5’-AAAGAATTCAGATCCACCTCCGCCGGTGAGGAAGGTGTTGGA-3’;

5’-TTTGAATTCGGCGGAGGTGGATCTCTGTACGGCCCACTGATG-3’ and 5’-AAACTCGAGCTAGACCTTGTACAGCGTCTGCAGCAGACT-3’. The amplified partial RanGAP1 DNA sequences corresponding to 1 to 514 amino acids were cloned into *Bam*HI / *Eco*RI sites of pcDNA4/V5 His ver. B vector, where DNA sequences of VC were cloned into *Hin*dIII /

*Bam*HI sites, followed by cloning of the partial RanGAP1 DNA sequences corresponding to 535 to 587 amino acids into *Eco*RI / *Xho*I sites to generate the construct of VC-Δ20aaRanGAP1. Then, the DNA sequences of VC-Δ20aaRANGAP1 were digested with *Hin*dIII / *Xho*I sites, and transferred into *Hin*dIII / *Xho*I sites in pMX vector.

2-2-2. Cell cultures and transfection

NIH3T3 cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS), 100 unit/mL Penicillin, and 100 μg/mL Streptomycin at 37°C under 5% CO_2. After the medium was exchanged to DMEM supplemented with 10% FBS, transfection by using a Lipofectamine 2000 Reagent (Invitrogen) was conducted according to the manufacturer's protocol. The NIH3T3 cells were subjected to subsequent analyses 24 hours after the transfection. To generate stable cell lines, selection of transfected NIH3T3 cells was performed by using a medium containing 500 μg/mL zeocin (Invitrogen).

2-2-3. Retrovirus production and infection

The plasmids for retrovirus production were introduced into PlatE cells by transfection. The transfected PlatE cells were grown at 37°C under 5% CO_2 for 8 hours. The medium that contained the retrovirus harboring the plasmid DNAs was collected 24 to 48 hours after the incubation and stocked at -30°C. NIH3T3 cells were incubated with a medium containing 8 μg/mL polybrene for 10 minutes before added with the medium stock that contained the retrovirus. After incubation for 24 hours, the infected NIH3T3 cells were subjected to subsequent FACS analysis.

2-2-4. FACS analysis and sorting

VN-SUMO2 stable cell lines infected with a retrovirus containing the VC-library DNAs were

treated with a trypsin, washed with PBS, and suspended in PBS. The same procedure was also applied to control cells for FACS analysis. The fluorescence intensities of 5,000 cells were examined by FACS (Epics Altra, BECKMAN COULTER) following the standard procedure. An excitation wavelength and a measurement wavelength were 488 nm and 525 (±15) nm light, respectively. The target cells with higher fluorescence intensity than the autofluorescence intensity of the control cells were sorted by FACS. The collected cells were seeded on a 6-well dish and incubated.

2-2-5. Identification of candidate SUMOylated proteins

The collected cells were seeded on a 10 cm dish with a low cell density to make the cells to be fully separated from one another. A few weeks after the cell incubation, each single cell clone was collected and incubated individually. The DNA of each single cell clone was extracted by a Wizard Genomic DNA Purification Kit (Promega). The VC-library DNA sequences were amplified by nested PCR using the following primers: 5'-CAAAGTAGACGGCATCGCAGC-3' and 5'-TTATGTATTTTTCCATGCCTTC-3' (primary PCR); 5'-TTTAAGCTTGCTAGCGCCATGAAGAACGGCATCAAGGCC-3' and 5'-TTATCGTCGACCACTGTGCTGGCGGCCGC-3' (secondary PCR). Each PCR product was subjected to electrophoresis through agarose gel that contained Ethidium Bromide for staining nucleic acids to check the amplification of the library DNA sequences. The detected single bands were extracted and purified using a FastGene Gel/PCR Extraction kit (GeneTics). In case of detection of several bands from one sample due to infection of multiple library DNAs to a single cell, individual DNA bands were collected by cutting the gel and purified in the same manner. After the purification of each PCR product, direct sequencing of the purified DNAs was conducted using the following primer: 5'-TGGTCCTGCTGGAGTTCGTG-3'. Homologies of the obtained sequencing data to nucleotide sequences in the GenBank database were analyzed using a BLAST search. The

amino acid sequences identical to the SUMO consensus recognition sites or the SUMO-Interacting Motifs were predicted by using the SUMOsp algorithm[1] and the SIM algorithm[2], respectively.

2-2-6. Confocal fluorescence microscopic analysis

The VN-SUMO2 stable cell lines were seeded on glass-based dishes and incubated at 37°C under 5% CO_2 for 24 hours. The cells were co-transfected with H2B-EBFP and either VC-RanGAP1 or VC-Δ20aaRanGAP1 and further incubated for 24 hours. After exchanging the medium to the observation buffer (DMEM modified with high glucose, L-Glutamine, HEPES, and no phenol red, supplemented with 10% FBS), the cells were observed using an IX81-FV-1000 confocal microscope (OLYMPUS Co. Ltd.) with the UPlanSApo 100×/1.40 oil objective. Excitation wavelength 405 nm and 515 nm were used for imaging EBFP and Venus, respectively. Images were analyzed using ImageJ software.

2-3. Results

2-3-1. Detection of SUMOylation by SUMO2 using reconstitution of the split Venus fragments.

Of the mammalian SUMO family members, SUMO2 was determined as a first target for the screening of SUMOylated proteins owing to its unique characteristics: polySUMO chain formation and stress responsiveness[3,4]. The N-terminal fragment of Venus, a bright yellow fluorescent protein derived from GFP[5], was genetically fused to the N-terminus of SUMO2 to generate VN-SUMO2 expression vector, connecting with a flexible GS linker (Gly-Gly-Gly-Gly-Ser) (Fig. 2-1). The C-terminal fragment of Venus was genetically fused to a protein that was encoded in mouse cDNA library to generate series of VC-library expression vectors. When the VC-library proteins are modified by VN-SUMO2 in living cells, both Venus fragments come close to each other. These

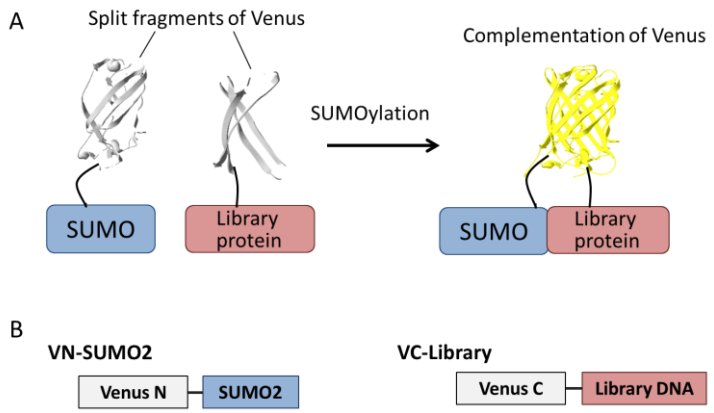

Fig. 2-1. The probes for detecting SUMOylation under a live-cell condition

(A) Schematic for detecting SUMOylation under a live-cell condition using the reconstitution of split Venus fragments. (B) Schematic structures of VN-SUMO2 and VC-library probes. Venus N (VN): N-terminal fragment (amino acids 1 to 158) of Venus. Venus C (VC): C-terminal fragment (amino acids 159 to 240) of Venus.

results in the reconstitution of Venus and recovery of fluorescence emission. The cells emitting fluorescence were screened and collected using FACS.

For the present system to detect SUMOylation, stable cell lines of mouse-derived NIH3T3 cells expressing VN-SUMO2 were generated. A famous SUMOylated protein RanGAP1[6–9] was used to confirm the recovery of Venus fluorescence upon SUMOylation of the substrate protein. Human RanGAP1 was connected with the C-terminal fragment of Venus to generate the VC-RanGAP1 expression vector. A SUMOylation site-deleted mutant of RanGAP1 that lacks 20 amino acids around the lysine 524 SUMOylation site[8] was also fused to the C-terminal fragment of Venus to generate VC-Δ20aaRanGAP1 expression vector for a negative control. The amino acids flanking the SUMOylation site was drastically deleted because it was reported that the amino acid sequences that flank the RanGAP1 SUMOylation site were recognized by SUMO modification machinery[10]. Confocal fluorescence microscopic analysis revealed that a high fluorescent intensity from reconstituted Venus was detected at the nuclear surface in cells that co-expressed VC-RanGAP1 and VN-SUMO2 (Fig. 2-2A). The detected localization of RanGAP1 was consistent with a previous report in which RanGAP1 translocated from the cytosol to the nuclear surface by SUMOylation[9].

Therefore, the localization of the detected reconstituted Venus suggested that RanGAP1 SUMOylation in living cells was appropriately visualized.

Subsequently, retrovirus infection of VC-RanGAP1 or VC-Δ20aaRanGAP1 cDNA into the VN-SUMO2 stable cell lines was performed, followed by FACS analysis of the fluorescence intensities (Fig. 2-2B). Approximately 43 (±3)% of the VC-RanGAP1-infected cells showed higher fluorescence intensity than the maximum fluorescence intensity of control cells (Fig. 2-2B, left). This indicated that the reconstituted Venus produced sufficient fluorescence intensity for FACS analysis to detect SUMOylation. Compared to control cells, the distribution of VC-Δ20aaRanGAP1-infected cells leaned relatively to higher fluorescence intensity region (Fig. 2-2B, right), which was due to false positives caused by high infection level. But, the number of the

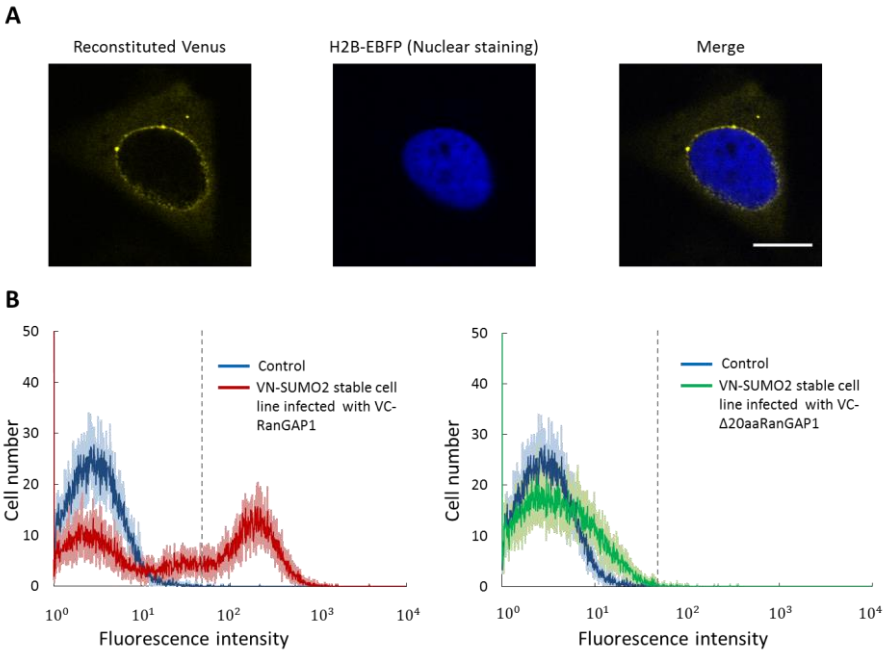

Fig. 2-2. Evaluation of the probes by using SUMOylated protein RanGAP1.

A. Fluorescence images of the VN-SUMO2 stable cell lines co-transfected with VC-RanGAP1 and H2B-EBFP. Scale bar: 10 μm. B. Fluorescence intensities of the non-infected VN-SUMO2 stable cell lines (control cells shown in blue) and the VN-SUMO2 stable cell lines infected with VC-RanGAP1 (left, shown in red) or VC-Δ20aaRanGAP1 (right, shown in green) were examined by FACS. Each histogram was acquired from measurements of 5,000 cells. The measurements were repeated 5 times. Light blue, light red, light green show the respective standard deviation. Dotted gray lines show the maximum florescence intensity of control cells where the cell number is almost zero.

VC-Δ20aaRanGAP1-infected cells that exceeded the maximum fluorescent intensity of the control cells was only 0.16 (±0.03)%. This indicates that the maximum fluorescence intensity of control cells could be used as a threshold to distinguish between SUMOylated proteins and non-SUMOylated proteins for the subsequent sorting. Taken together, it was concluded that the SUMOylation of target proteins was detectable by the reconstitution of the Venus fragments in the present method.

2-3-2. Identification of the candidate SUMOylated proteins.

The present method was applied to screen mammalian SUMOylated protein candidates (Fig. 2-3). The VN-SUMO2 stable NIH3T3 cell lines were infected with the retrovirus containing cDNA libraries that were attached with the cDNA of VC fragment (VC-library). The infection efficiency was modulated to up to 30% so that a single VC-library DNA was transferred into each cell. The percentage was calculated, which was estimated using FACS analysis of NIH3T3 cells infected with a GFP expression retrovirus vector produced under the same procedure with the VC-libraries (Fig. 2-4). A few days after the infection, some of the infected cells showed higher fluorescent intensities than that of the control cells in the FACS analysis (Fig. 2-4A). This result suggested that modification of VN-SUMO2 to some of the VC-library proteins took place, resulting in subsequent reconstitution of Venus and fluorescence emission. The cells emitting fluorescence were sorted by FACS and incubated for a week to make the cells proliferate. Subsequently, sorting of the fluorescent cells was conducted again to accurately separate the target fluorescent cells from the non-fluorescent cells. The process of sorting and incubation was repeated 3-4 times. FACS analysis of the finally-obtained cell population showed that the fluorescent cells were successfully isolated (Fig. 2-5). Then, the individual fluorescent cells were separately seeded on a culture dish, followed by incubation for a few weeks and isolation of the single-cell clones. From the DNA sequence analysis

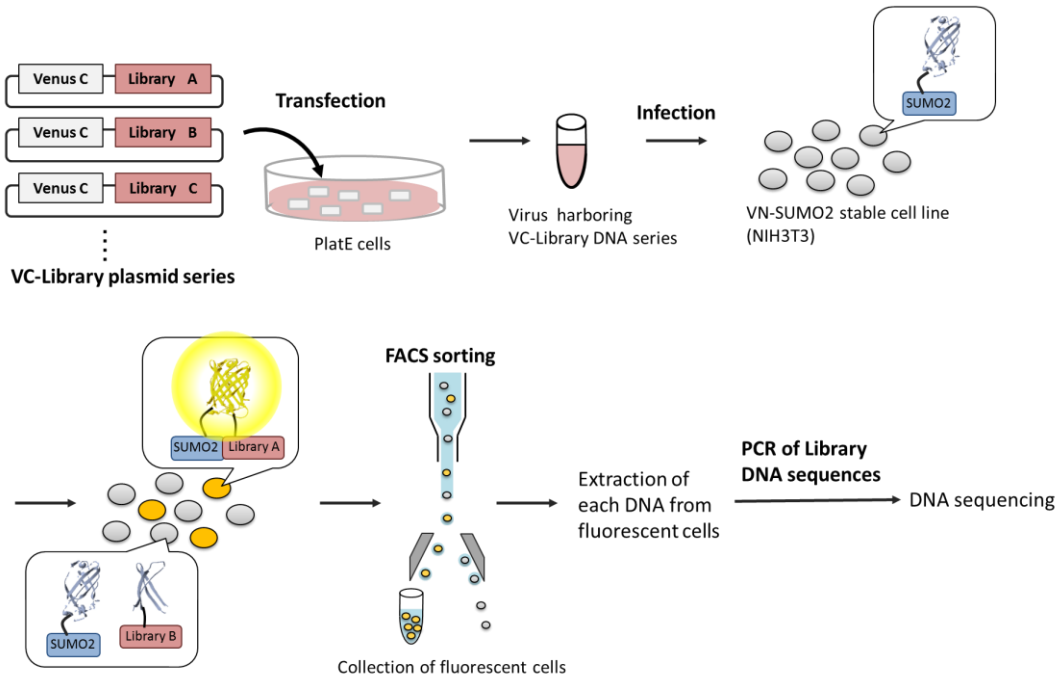

Fig. 2-3. Schematic of screening mammalian SUMOylated proteins based on the reconstitution of split Venus fragments.

The virus infection vectors that harbor library DNAs with VC DNA are introduced into PlatE cells. The produced viruses containing VC-library DNAs are added to NIH3T3 cells that stably express VN-SUMO2. The fluorescent cells containing reconstituted Venus are sorted by FACS. Each library DNA is extracted from each fluorescent cell. Candidates of SUMOylated proteins are identified by an analysis of the library DNA sequences.

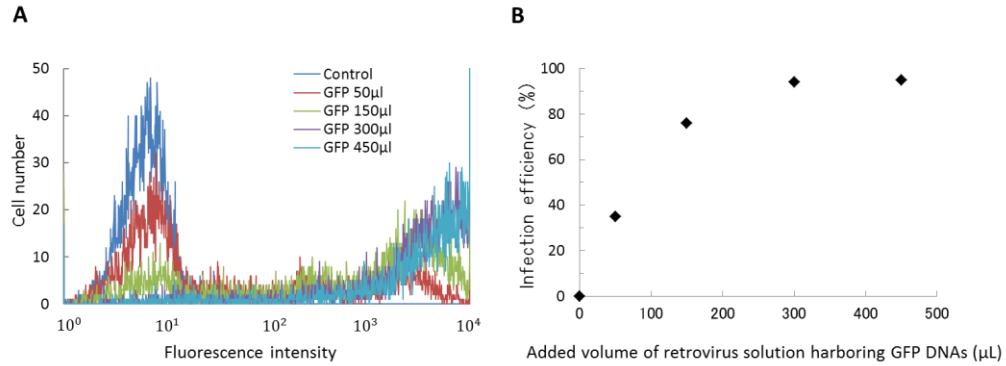

Fig. 2-4. Control infection with retroviruses harboring GFP DNAs.

(A) Fluorescence intensities of the VN-SUMO2 stable cell lines with or without addition of retrovirus harboring GFP DNAs. Different volume of the retrovirus solutions was added as indicated. (B) The infection efficiency of Fig. 2-4A was calculated by the ratio of the cells with higher fluorescent intensities than those of the control cells to total cells. The number of total cells was 5,000.

of the cDNAs that were extracted from the isolated single cell clones, 38 SUMOylated protein candidates were identified (Table 2-1). Among the identified candidates, 17 proteins contained the amino acid sequences identical to SUMO consensus recognition site shown as Ψ-K-X-E/D ("Ψ": a hydrophobic amino acid, "K": the SUMO-acceptor lysine residue, "X": one of any amino acids, "E": a glutamic acid, "D": an aspartic acid) (Table 2-2). According to previous reports, the screened candidate proteins show various localization in intracellular compartments: Anxa5 in the cytoplasm and the nucleus[11], Drosha in the nucleus[12], Plscr3 in mitochondria[13] , and Tuba1b in microtubules[14]. Also, the identified candidate proteins show diverse cellular functions: Narf is an ubiquitin ligase[15], Myof regulates membrane integrity to affect vascular endothelial growth factor signaling[16], and Arpc1b is related to cell cycle[17], and Taz is implicated in cell proliferation[18]. This diversity indicates that the present method allows for screening various SUMOylated protein candidates with no bias in terms of their normal intracellular location and function in living mammalian cells. Of the candidates, two proteins, Rpl37a and Lmna, have already been reported to be SUMOylated by SUMO2 in

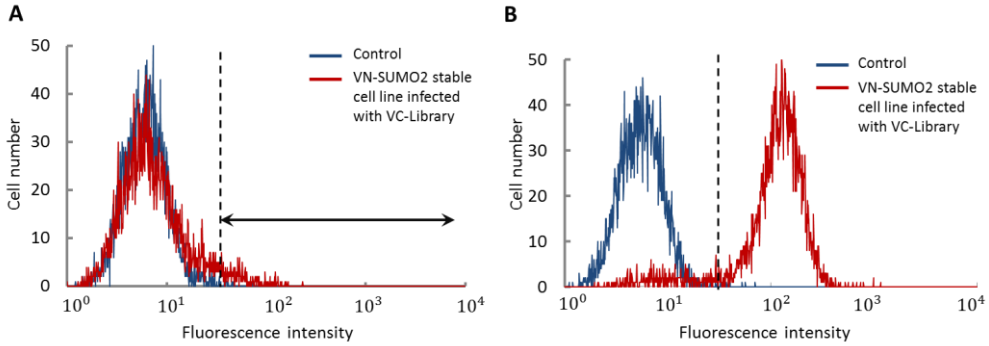

Fig. 2-5. FACS sorting of the fluorescent cells.

FACS analysis of fluorescence intensities of VN-SUMO2 stable cell lines with or without infection of VC-library DNAs. (A) Comparison of control cells with the library-DNA-infected cells before FACS sorting. The region shown as an arrow contains the cells with higher fluorescence intensities than control cells. (B) Comparison of control cells with the library-DNA-infected cells after FACS sorting. The fluorescent cells included in the target region were repeatedly incubated and sorted by a FACS four times. The data show the fluorescence intensity of the finally-obtained cells.

mammalian cells[19,20]. This confirms that the present screening method can identify SUMOylated proteins using living mammalian cells. However, the present method also has a potential to detect proteins that non-covalently interact with SUMOs via SUMO-interacting motifs (SIMs) because the Venus fragments can reconstitute by coming close to each other. The amino acid sequences of the candidate proteins were checked whether the identical sequences to SIMs are included or not (Table. 2-2). 24 candidate proteins contained the same amino acid sequences with SIMs, indicating that non-SUMOylated proteins are possibly included among the detected candidate proteins. Though further biochemical analysis, such as Western blotting, is required to confirm SUMOylation of the candidate proteins, the present screening method could detect 2 SUMOylated proteins and identify 36 SUMOylation-unreported proteins as candidates for novel SUMOylated proteins without limitations in subcellular protein locations and functions.

Table. 2-1. The candidates of SUMOylated proteins identified by analyzing library DNA sequences of the sorted fluorescent cells.

Classification	SUMOylated protein candidates
Reported as SUMOylated	Rpl37a, Lmna
Not reported as SUMOylated	Rps9, Rpl32, Eif3e, Gsn, Stx12, Bgn, Drosha , Uqcrh, Plxnb2, Rpl18a, Atac2, Ermp1, Mrpl4, Tmsb4x, Rpsa, Lgals3, Pcolce, Tuba1b, Pbrm1, Myof, Dynlrb1, Fam63b, Taz, Rps3a, Myl9, Rpl6, Narf, Arpc1b, Psmb4, Polr1d, Rpl10, Fth1, Anxa5, Plscr3, Wisp2, Cops7a

Table. 2-2. Classification of the identified candidates of SUMOylated proteins based on the presence of SUMO consensus recognition site or SUMO-interacting motif (SIM).

Type of the included sequences	SUMOylated protein candidates
SUMO consensus recognition site	Lmna, Rps9, Drosha , Uqcrh, Plxnb2, Rpl18a, Atac2, Ermp1, Rpsa, Tuba1b, Pbrm1, Myof, Fam63b, Rps3a, Narf, Psmb4, Anxa5
SUMO-interacting motif (SIM)	Lmna, Eif3e, Gsn, Stx12, Bgn, Drosha , Plxnb2, Atac2, Ermp1, Mrpl4, Rpsa, Lgals3, Pcolce, Tuba1b, Pbrm1, Myof, Dynlrb1, Fam63b, Rpl6, Narf, Arpc1b, Rpl10, Anxa5, Cops7a

2-4. Discussion

In this study, reconstitution of split fluorescent protein fragments was used for screening of mammalian SUMOylated proteins. This technique has several beneficial points in the screening. Conventionally, IP-based MS analytical screening methods have been widely used to comprehensively detect SUMOylated proteins[21,22] and further identify SUMOylation sites in various biological contexts[23–26]. The versatile analyses by IP-based MS methods are very useful to obtain fundamental insights into the SUMOylation roles in cells. The IP-based MS analytical methods required the enzymatic digestion of the proteins collected from cell lysates in the sample preparation process. Because the deSUMOylation by SUMO proteases took place not only in living cells but also potentially after cell lysis, the possibility of detecting SUMOylated proteins could be decreased in the IP-based MS analytical screening approaches. In contrast, the fluorescent signals produced by the reconstitution of split fluorescent protein fragments maintain even though deSUMOylation occurs, due to the irreversibility of the reconstitution reaction. Hence, the present method would have a potential in detecting infrequently-SUMOylated proteins. The single cell analysis by FACS would also increase the possibility of detecting rarely-SUMOylated proteins. In the present method, each fluorescence signal from the cells co-expressing VN-SUMO2 and VC-Library was detected by FACS. Hence, the weak SUMOylation signal in a cell could be potentially detected without being

buried in the strong SUMOylation signal in another cell. In the IP-based MS method, frequently-SUMOylated proteins are preferentially collected from cell lysates by immunoprecipitation. Thus, the detectable scope of SUMOylated proteins was potentially biased. In contrast, in the present method, both library proteins and SUMO proteins were exogenously expressed and the fluorescent signals were respectively detected in the individual cells by FACS. Therefore, the bias of the detectable proteins due to the differences of SUMOylation level upon protein species could be lowered in the present method. Compared to the dataset of the detected proteins by the previous IP-based MS analytical methods[27], 24 proteins of the 38 proteins detected by the present method were also screened, which supported the reliability of the present method for screening SUMOylated proteins. Whereas, the rest 14 proteins were not detected by the IP-based MS analytical screening, which indicated that the present approach using the reconstitution of the split fluorescence protein fragments had a potential to enable to partially detect different targets from the proteins detected by the previous MS-based approaches. Though the detectable targets were not so many at the present stage, enlarging the cDNA libraries would increase the number of the detectable proteins. Taken altogether, the present system using the reconstitution of split fluorescence protein fragments has unique characteristics in detecting SUMOylation, which could contribute to the complemental detection for wide ranges of SUMOylated proteins.

Another advantage is that the range of detectable candidate proteins is not limited by the proteins' subcellular localization. The previous screening method based on the yeast two-hybrid method[28] required that the target proteins move into the nucleus, where SUMOylation triggered transcription of the reporter gene. Therefore, it is difficult to examine SUMOylation that occurs in other intracellular compartments, such as ER or Golgi apparatus, by this method. In contrast, the present screening method can evaluate the SUMOylation that takes place in any intracellular locations because the split fluorescent protein fragments can reconstitute anywhere in the cell. In the present

research, various candidate proteins localized in distinct subcellular localization, such as Anxa5 in the cytoplasm and nucleus[11], Drosha in the nucleus[12], Plscr3 in mitochondria[13], and Tuba1b in microtubules[14], were detected. These results supports that proteins diversely-compartmentalized in cells can be targeted for assessing SUMOylation by the present method.

Recently, a screening method for yeast SUMOylated proteins based on split fluorescent protein reconstitution has been reported[29]. The method using yeast cells has several difficulties in investigating mammalian SUMOylation because the complexities of the mammalian SUMOylation system, including the specificity of SUMO isoforms and the SUMOylation patterns dependent on cell types, cannot be reflected in yeast cells. In contrast, the present method can be applied to investigate both SUMO-isoform-specific and cell-type-specific SUMOylation by replacing SUMO isoforms or target cell lines. For one example, though murine NIH3T3 cells were used in the present study, human-derived cell lines can be also applied by using the retrovirus that can infect the human cells. Moreover, the present screening method would be easily applicable to looking into SUMOylation that respond to stresses such as UV irradiation or heat shock[4]. Hence, the present method is adequate for inspecting mammalian SUMOylation under different cellular conditions.

The present method enabled to detect SUMOylation by the reconstitution of split Venus fragments. However, some SUMOylated proteins would not be assessed due to the problems in preparing cDNA libraries. First, linker length between the library proteins and the Venus fragments might be inappropriate for some proteins due to the difference in the relative location between the amino acid termini and SUMOylation sites. In the present study, the library proteins and the Venus fragments were fused with GS linker (Gly-Gly-Gly-Gly-Ser). By using the linker length, 36 proteins were detected as SUMOylated protein candidates. However, some library proteins might be not detected because the linker length was too short to make the Venus fragments come close to each other. Second, owing to the protocol to prepare cDNA libraries, full-length DNA sequences coding

library proteins were difficult to be inserted into the expression plasmids, which may cause loss of their SUMOylation sites. In the present preparation method, the library DNAs were digested with two restriction enzymes (*Eco*RI and *Not*I sites) and transferred to the plasmids containing Venus fragment (See Materials and Methods). Hence, DNA sequences harboring the restriction sites were partially inserted into the obtained plasmids. The reason that E1/E2/E3 enzymes were not detected in the present screening was possibly due to the problems of preparing cDNA libraries. To solve the problems, the cDNA library preparation protocol should be altered in future; preparation of various length of GS-linkers; application of other cloning techniques such as recombinase, which does not require restriction enzymes, to fuse DNAs encoding full-length proteins to Venus fragments.

Besides, sorting accuracy is significant for identifying the SUMOylated proteins. In this study, an old-type FACS machine was used. It requires repetition process of cell incubation and cell sorting to decrease false-positive rates. This might result in detecting overlapped SUMOylated protein candidates derived from same single clones, limiting a diversity of detectable protein species. This issue will be solved by latest FACS machines with superior sorting accuracy, which would reduce the number of the repetition of cell incubation and cell sorting. In addition, total time required for the sorting process will be considerably shortened if the latest cell sorter is available.

In conclusion, I developed a new system for screening mammalian SUMOylated proteins using living cells. The present screening method was based on reconstitution of split fluorescent protein fragments and FACS sorting. The fluorescence protein reconstitution is prospective for detecting infrequently-SUMOylated proteins due to the characteristics: no requirement of a cell-disruption process for the detection of SUMOylation; maintenance of SUMOylation signals as fluorescence owing to the irreversible reconstitution reaction; no limitation in subcellular locations of target proteins to be SUMOylated. By the present screening method, 38 proteins with diverse subcellular functions and localizations were identified. Of the detected proteins, 2 proteins have been already

reported to be SUMOylated proteins. The other 36 proteins have not been confirmed to be SUMOylated, prospective candidates for novel SUMOylated proteins. The present method detected partially different proteins from the proteins detected by conventional IP-based MS screening methods. Therefore, the present screening method has a potential to complementarily explore SUMOylated proteins in a different manner from the conventional methods.

Reference

1. Xue, Y., Zhou, F., Fu, C., Xu, Y. & Yao, X. SUMOsp: a web server for sumoylation site prediction. *Nucleic Acids Res.* **34,** W254–W257 (2006).
2. Zhao, Q. *et al.* GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res.* **42,** W325-30 (2014).
3. Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–82 (2004).
4. Tempé, D., Piechaczyk, M. & Bossis, G. SUMO under stress. *Biochem. Soc. Trans.* **36,** 874–878 (2008).
5. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
6. Matunis, M. J., Coutavas, E. & Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **135,** 1457–1470 (1996).
7. Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88,** 97–107 (1997).
8. Macauley, M. S. *et al.* Structural and dynamic independence of isopeptide-linked RanGAP1 and SUMO-1. *J. Biol. Chem.* **279,** 49131–49137 (2004).
9. Matunis, M. J., Wu, J. & Blobel, G. SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *J. Cell Biol.* **140,** 499–509 (1998).
10. Rodriguez, M. S., Dargemont, C. & Hay, R. T. SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J. Biol. Chem.* **276,** 12654–12659 (2001).

11. Sun, J., Salem, H. H. & Bird, P. Nucleolar and cytoplasmic localization of annexin V. *FEBS Lett.* **314,** 425–429 (1992).

12. Tang, X., Zhang, Y., Tucker, L. & Ramratnam, B. Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. *Nucleic Acids Res.* **38,** 6610–6619 (2010).

13. Liu, J., Chen, J., Dai, Q. & Lee, R. M. Phospholipid scramblase 3 is the mitochondrial target of protein kinase C delta-induced apoptosis. *Cancer Res.* **63,** 1153–1156 (2003).

14. Azakir, B. A., Fulvio, S. Di, Therrien, C. & Sinnreich, M. Dysferlin interacts with tubulin and microtubules in mouse skeletal muscle. *PLoS One* **5,** 1–10 (2010).

15. Yamada, M. *et al.* NARF, an nemo-like kinase (NLK)-associated ring finger protein regulates the ubiquitylation and degradation of T cell factor/lymphoid enhancer factor (TCF/LEF). *J. Biol. Chem.* **281,** 20749–20760 (2006).

16. Bernatchez, P. N. *et al.* Myoferlin regulates vascular endothelial growth factor receptor-2 stability and function. *J. Biol. Chem.* **282,** 30745–30753 (2007).

17. Molli, P. R. *et al.* Arpc1b, a centrosomal protein, is both an activator and substrate of Aurora A. *J. Cell Biol.* **190,** 101–114 (2010).

18. Lei, Q. *et al.* TAZ Promotes Cell Proliferation and Epithelial-Mesenchymal Transition and Is Inhibited by the Hippo Pathway. *Mol. Cell. Biol.* **28,** 2426–2436 (2008).

19. Yun, C. *et al.* Nucleolar protein B23/nucleophosmin regulates the vertebrate SUMO pathway through SENP3 and SENP5 proteases. *J. Cell Biol.* **183,** 589–595 (2008).

20. Zhang, Y.-Q. & Sarge, K. D. Sumoylation regulates lamin A function and is lost in lamin A mutants associated with familial cardiomyopathies. *J. Cell Biol.* **182,** 35–39 (2008).

21. Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K. & White, M. A. Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J. Biol. Chem.* **279,** 20999–21002 (2004).

22. Tirard, M. *et al.* In vivo localization and identification of SUMOylated proteins in the brain of His6-HA-SUMO1 knock-in mice. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21122–21127 (2012).

23. Lamoliatte, F. *et al.* Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nat. Commun.* **5,** 5409 (2014).

24. Hendriks, I. A. *et al.* Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol.* **24,** 325–336 (2017).

25. Lamoliatte, F., McManus, F. P., Maarifi, G., Chelbi-Alix, M. K. & Thibault, P. Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification. *Nat. Commun.* **8,** 14109 (2017).

26. Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **21,** 927–936 (2014).

27. Hendriks, I. A. & Vertegaal, A. C. O. A comprehensive compilation of SUMO proteomics. *Nat. Rev. Mol. Cell Biol.* **17,** 581–595 (2016).

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

29. Sung, M.-K. *et al.* Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. *Genome Res.* **23,** 736–746 (2013).

Chapter 3.

Confirmation of SUMOylation of the detected candidate proteins

3-1. Introduction

I developed a new screening system for mammalian SUMOylated proteins using combination of reconstitution of split Venus fragments and FACS sorting. By the present screening method, 36 proteins were identified as novel SUMOylated protein candidates. However, the possibility of false positives could not be denied. Therefore, I performed further analysis to confirm SUMO2 modification to these candidates, using immunoprecipitation and Western blotting methods.

3-2. Materials and Methods

3-2-1. Construction of expression plasmids

From human genome DNA of Hela cells, the DNA fragments of Myc-SUMO-1 were amplified by PCR using the following primers: 5'-GGCGCTAGCACCATGGAACAAAAACTCATCTCAGAAGAGGATCTGAAGCTTATGTCTGACCAGGAGGCAAAACCTT-3' and 5'-TTTCTCGAGCTAAACTGTTGAATGACCCCCCGTTTGTTC-3', and then cloned into *Nhe*I / *Xho*I sites of pcDNA4/V5 His ver. B vector. Similarly, the DNA fragments of Myc-SUMO2 were amplified by PCR using the following primers: 5'-GGCGCTAGCACCATGGAACAAAAACTCATCTCAGAAGAGGATCTGAAGCTTATGGCCGACGAAAAGCCCAAG-3' and 5'-TTTCTCGAGCTAGTAGACACCTCCCGTCTGCTGTTGGAA-3', and cloned into *Nhe*I / *Xho*I sites of pcDNA4/V5 His ver. B vector. The fragments of Venus were amplified by PCR using the following primers: 5'-TTTGCTAGCATGGTGAGCAAGGGCGAGGAGCTGTTCAC-3' and 5'-TTTGCTAGCTCCGCCTCCGCCCTTGTACAGCTCGTCCA-3', and cloned into *Nhe*I site of the construct of Myc-SUMO1 or Myc-SUMO2 to generate the constructs of Venus-Myc-SUMO1 and Venus-Myc-SUMO2, respectively. The cDNA encoding Atac2 was amplified by nested PCR

from total cDNAs of mouse brain (GenoStaff) using the following primers: 5’-AAGGTCTGGGGGATCCCGGCAGCTTAGGAGTGCGTGTGAGCAGG-3’ and 5’-CACAGGAGACCAGGGCCCACTTCGAGGATAAGGGCCTGCAGGGC-3’ (primary PCR); 5’-TTTGAATTCATGGATAGTAGCATCCACCTGAGTGGGCTCCTCAG-3’ and 5’-TTTCTCGAGTCTAGACGGCGTCGGAGCCTCAGGAAGAATGCATG-3’ (second PCR). Before the second PCR, the primary PCR product was subjected to agarose gel electrophoresis, and the gel around the target size was cut and purified. The purified primary PCR product was used as a template for the second PCR. The cDNA encoding Plscr3 was amplified by PCR from total cDNAs of mouse brain (GenoStaff) using the following primers: 5’-TTTAAGCTTATGGCAGGCTACTTGCCCCCCAAAGGCTATGCCC-3’ and 5’-TTTCTCGAGCGACTGGT

GATGGCAGAGGGTCCTGCGCCTCCTC-3’. The amplified fragments were digested with *Hin*dIII / *Xho*I and then cloned into *Hin*dIII / *Xho*I sites of pcDNA4/V5 His ver. B vector to generate the construct of Plscr3-V5. The DNA of N-terminally Flag-fused Atac2 was amplified by PCR using the following primers: 5’-TTTGAATTCCGACTACAAGGACGATGACGACAAGGGCGATAGTA-3’ and 5’-TTTGCGGCCGCTCAGCGTCGGAGCCTCAGGAAGAATGCATGCTT-3’, and then cloned into *Eco*RI / *Not*I of pTriEx-3 vector. K305A-V5, K408A-V5, K749A-V5, K408R-V5, Flag-K408R, and SUMO2 K11R were generated by overlapping PCR. The primers used for the point mutations were as follows: K305A, 5’-TCTGATTTTGGAGGCAGGAGAAGTGATCG-3’ and 5’-CGATCACTTCTCCTGCCTCCAAAATCAGA-3’ ; K408A, 5’-CCGGAACAGATAGCACAGGAAGTGGAC-3’ and 5’-GTCCACTTCCTGTGCTATCTGTTCCGG-3’ ; K749A, 5’-GTTTGGCTTCGCGACCGAGGAGTAT-3’ and 5’- ATACTCCTCGGTCGCGAAGCCAAAC-3’ ;

K408R, 5’- CGGAACAGATAAGACAGGAAGTGGA-3’ and 5’-TCCACTTCCTGTCTTATCTGTTCCG-3’; SUMO2 K11R, 5’-GGAGTCAgGACTGAGAACAACGATCA-3’ and 5’-CTCAGTCcTGACTCCTTCCTTGGGCT-3’. The fragments of Venus were amplified by PCR using the following primers: 5’-TTTGCTAGCATGGTGAGCAAGGGCG-3’ and 5’-TTTGAATTCTCCGCCTCCGCCCTTGTACAGCTCGTCCATGCCG-3’, and then cloned into *Nhe*I / *Eco*RI sites of Atac2-V5 to generate the construct of Venus-Atac2. The fragments of VC were amplified by PCR using the following primers: 5’-TTTGCTAGCATGAAGAACGGCAT-3’ and 5’-TTTGAATTCTCCGCCTCCGCCCTTGTAC-3’, and then cloned into *Nhe*I / *Eco*RI sites of Atac2-V5 to generate the construct of VC-Atac2. The fragments of N-terminally-V5-fused Atac2 were amplified by PCR using the following primers: 5’-TTTGCTAGCATGGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGGATAGTAGCATCCACCTGAG-3’ and 5’-TTTGAATTCCGGCGTCGGAGCCTCAGGAAGAATGCATG-3’, and then cloned into *Nhe*I / *Eco*RI sites of pcDNA4 V5 His B vector to generate V5-Atac2-V5. His-SUMO2 construct was generated as previously described[1]. GST-E1 and His-E2 constructs were generated as previously described[2].

3-2-2. Cell cultures and transfection

NIH3T3 cells and HEK293T cells were cultured in the same manner described in the Chapter 2. Transfection of NIH3T3 cells was performed in the same procedure described in the Chapter 2. HEK293T cells were transfected by using PEI MAX (Polyscience Inc.) according to the manufacturer’s protocol. After 24 hours from the transfection, the cells were subjected to subsequent analyses.

3-2-3. Immunoprecipitation and Western blotting

The method for immunoprecipitation of SUMOylated proteins was described in the previous report[3]. In brief, the transfected cells were washed with PBS, harvested by scraping in 1 ml PBS, and transferred to a 1.5 ml microcentrifuge tube. The cells were collected by centrifugation at 13,000 g for 30 seconds, and the supernatant was removed. The collected cells were lysed in 1 ml NP-40 lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40) supplemented with 20 mM N-ethylmaleimide (freshly dissolved) and complete protease inhibitor (Roche). The lysate was centrifuged at 16,000 g for 10 minutes at 4°C, and the supernatant was transferred to another microcentrifuge tube to remove cellular debris. To the collected supernatant, 0.8-1.0 μL of anti-V5 antibody (Invitrogen) was added, and the sample tube was rotated at 4°C for 1 hour. The sample tube was then again rotated at 4°C for 2 hours after appropriate amount of protein G-Sepharose (GE Healthcare) was added to the tube. The protein G-Sepharose beads were washed with the lysis buffer 5 times. The beads were collected by centrifugation at 16,000 g for 10 seconds, 4°C after each wash. To the beads, 50 μL of 2 × SDS sample buffer (25 mM Tris-HCl, pH6.8, 10% 2-mercaptoethanol, 4% SDS, 10% Sucrose, 0.000004% Bromophenol blue) was added, and the sample was boiled at 95°C for 5 minutes.

The immunoprecipitated proteins by the above procedure were subjected to SDS-PAGE and transferred to nitrocellulose membrane (Amersham™ Hybond™-ECL, GE Healthcare). Blocking of the membranes was performed with 1% skim milk (BD) in TBS-T buffer (50 mM Tris-HCl, pH8, 150 mM NaCl, 0.05% Tween-20). The primary antibodies were as follows: mouse monoclonal anti-V5 antibody (Invitrogen) diluted 1:5000 to 1:10000; rabbit polyclonal anti-Myc antibody (MBL) diluted 1:1000; and mouse monoclonal anti-FLAG antibody (Sigma) diluted 1:1000. The primary antibodies were detected by proper secondary antibodies, anti-mouse IgG or anti-rabbit IgG, labeled with horseradish peroxidase (Invitrogen). The immunoblot bands were visualized using

SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific), with image analyzers LAS1000 plus (Fuji photo Film Co. Ltd.) or LAS4000 mini (GE Healthcare).

3-2-4. Confocal fluorescence microscopic analysis

Basically, the procedure is same with that described in the Chapter 2. Briefly, the transfected cells seeded on glass-based dishes were observed using an IX81-FV-1000 confocal microscope (OLYMPUS Co. Ltd.) with a UPlanSApo 100×/1.40 oil objective. Excitation wavelength 405 nm and 515 nm were used for imaging EBFP and Venus, respectively. Images were analyzed using ImageJ software.

3-2-5. *In vitro* SUMOylation assay of recombinant Atac2 proteins

Recombinant FLAG-Atac2 or FLAG-K408R proteins were expressed in HEK293T cells and purified by batch absorption using Anti-Flag M2 Agarose Affinity Gel (Sigma) according to the manufacturer's protocol. The recombinant proteins binding to the column were eluted by competition with 100 μg/ml FLAG peptide (Sigma) in TBS buffer (10 mM Tris-HCl, 150 mM NaCl, pH 7.4). The collected proteins were concentrated using Vivaspin 20-50K (GE Healthcare) and dissolved in the SUMOylation assay buffer (50 mM Tris-HCl [pH7.6], 6 mM $MgCl_2$, 1 mM DTT). The recombinant His-SUMO2, GST-E1, and His-E2 proteins were purified as previously described[1,4]. Incubation of 2.5 μg of Flag-Atac2 or Flag-K408R proteins was performed for 2 hours at 30 ℃ in the absence or presence of 2 mM ATP (Sigma), 1.0 μg of His-SUMO2, 0.75 μg of GST-E1, and 0.05 μg of His-E2. The mixtures were subjected to subsequent Western blotting analysis.

3-3. Results

3-3-1. Assessment of SUMOylation of the candidate proteins.

To confirm SUMO2 modification of the candidate proteins, the conventional analysis using immunoprecipitation and Western blotting was performed. The candidate proteins and SUMO2 were genetically fused with V5 and Myc epitope tags, respectively. Lysates of NIH3T3 cells expressing V5-tagged candidate proteins with or without Myc-SUMO2 co-expression were immunoprecipitated with anti-V5 antibodies. The immunoprecipitated proteins were denatured and subjected to SDS-PAGE, a gel electrophoresis step to separate proteins based on the difference of the molecular weight. In the denaturation condition, a covalent isopeptide bond formed between a SUMO and a protein is not cleaved. On the other hand, non-covalent protein-protein interaction is lost. Thus, through the gel electrophoresis, SUMO2 modification can be distinguished from non-covalent interaction by the upshift in the molecular weight. The proteins in the gel were transferred onto a membrane, and V5-tagged candidate proteins and Myc-tagged SUMO2 proteins were detected by Western blotting analysis with anti-V5 antibodies and anti-Myc antibodies, respectively. If V5-tagged candidate proteins are modified by Myc-SUMO2 proteins, the detected size by immunoblot with anti-V5 antibody should be consistent with the size detected by immunoblot with anti-Myc antibody.

When immunoprecipitates from cells co-expressing one of the candidate proteins Plscr3-V5 (35 kDa) and Myc-SUMO2 (12 kDa) were blotted with anti-Myc antibodies, Myc-SUMO2 proteins were multiply detected in different sizes (Fig. 3-1). However, co-existence of Plscr3-V5 proteins was not clearly detected in the same sizes with those of the detected multiple Myc-SUMO2 proteins by immunoblot with anti-V5 antibodies. From its amino acid sequences, Plscr3 doesn’t have SUMO consensus recognition sites but contain SUMO-interacting motifs (SIMs) (Chapter 2, Table 2). This

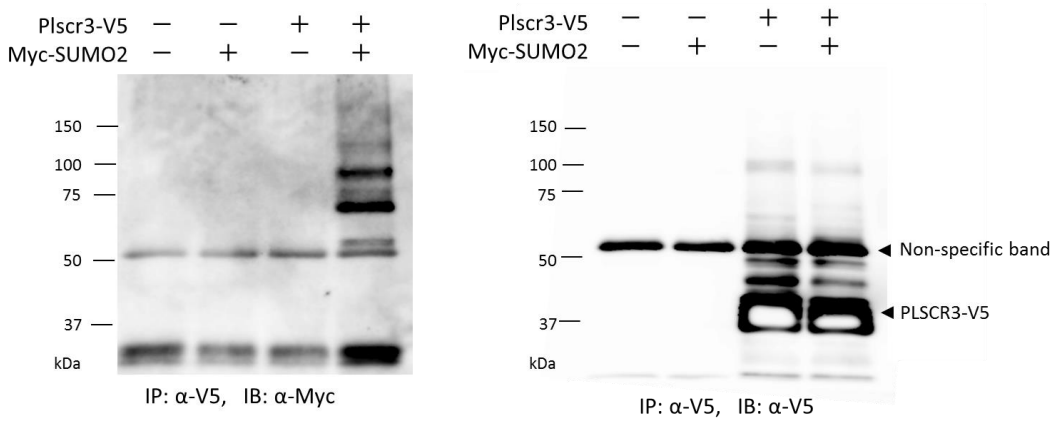

Fig. 3-1. Assessment of SUMOylation of Plscr3.

NIH3T3 cells were transfected with the expression plasmids as indicated. Immunoprecipitates with anti-V5 antibodies from each transfected cells were subjected to Western blotting with anti-Myc antibodies (Left) and anti-V5 antibodies (Right).

indicates the possibility that Plscr3 non-covalently interact with other SUMOylated proteins via SIMs.

When immunoprecipitates from cells co-expressing another candidate protein Atac2-V5 (92 kDa) and Myc-SUMO2 were analyzed, Myc-SUMO2 proteins were detected in approximately 120 kDa and over 150 kDa by immunoblot with anti-Myc antibodies. Atac2-V5 proteins were detected in approx. 90 kDa, whose size is consistent with the size of Atac2-V5, and in 120 kDa, whose size is also blotted with anti-Myc antibody, by immunoblot with anti-V5 antibodies (Fig. 3-2). The size difference (approx. 30kDa) from unmodified Atac2-V5 was slightly larger than the size of Myc-SUMO2 (12kDa). In previous reports, SUMO-modified target proteins reduced the mobility in the gel[5–7]. Hence, the detected upshifted band can be rationally assigned to the SUMOylated Atac2. On the other hand, the co-existence of Atac2-V5 proteins was not detected in the same sizes as Myc-SUMO2 proteins that were detected over 150 kDa in a ladder-like manner. It is unclear whether the Myc-SUMO2 proteins over 150 kDa were originated from Atac2 modified by either multiple SUMO proteins or other post-translational protein modifiers. It is also possible that other SUMOylated proteins were co-precipitated with SUMOylated Atac2. The origin of the upshifted

bands over 150 kDa was discussed in a later section (Examination of the origin of the unknown 150kDa Myc-SUMO2 band).

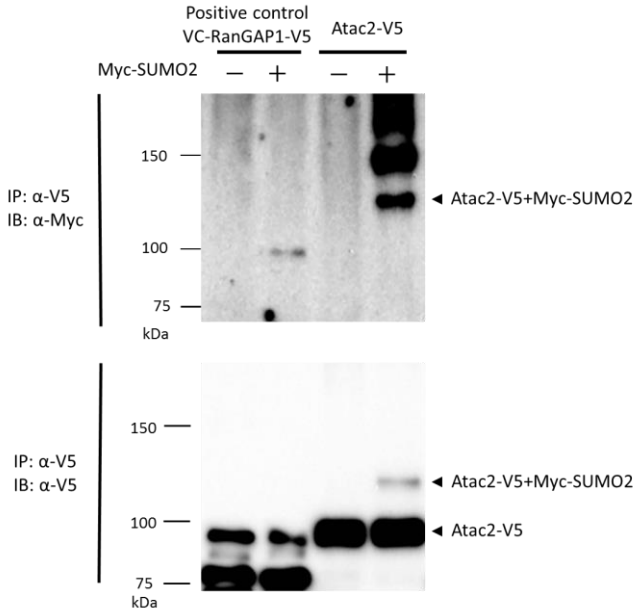

Fig. 3-2 Atac2 is SUMOylated.

NIH3T3 cells were transfected with V5-tagged Atac2 in the absence or presence of Myc-tagged SUMO2 expression, subjected to immunoprecipitated with anti-V5 antibodies. The precipitated proteins were resolved by SDS-PAGE and analyzed by Western blot with anti-Myc antibodies (Top) and anti-V5 antibodies (Bottom). As a positive control, V5-tagged VC-RanGAP1 was shown.

3-3-2. K408 is the SUMOylation site in Atac2.

Next, to identify the SUMOylation site in Atac2, putative SUMO-acceptor lysine residues were predicted by SUMO·sp version 1.0 online web server (http://sumosp.biocuckoo.org/1.0/) that was based on the algorithm of searching for SUMO consensus recognition sites Ψ-K-X-E/D[8]. Three lysine residues, K305, K408, and K749, were selected as the candidates of SUMOylation sites in Atac2 (Table 3-1). Three mutants, in which one of the predicted lysine was substituted to an alanine, were genetically prepared. The lysine residues were changed to alanine residues. Cells that expressed

wild-type Atac2 or the mutated Atac2 (K305A, K408A, and K749A) with or without Myc-SUMO2 co-expression were subjected to the immunoprecipitation and Western blotting analysis in the same manner as that described in the section of "assessment of SUMOylation". In K305A or K749A mutants, SUMOylation bands were still detected, as was the case of the wild type Atac2 (Fig. 3-3). In contrast, the K408A mutant did not show any Myc-SUMO2 band, indicating that K408A mutant could not be SUMOylated. To exclude the possibility that the change of surface charge upon K to A mutation inhibited the SUMOylation of Atac2, a mutant where K408 was mutated to an arginine residue in order not to change the net surface charge was prepared, and its SUMOylation was assessed. The result showed that no SUMOylation bands were detected in K408R mutant (Fig.3-4). Taken all together, it was concluded that K408 was the SUMOylation site in Atac2.

Table 3-1. Potential SUMOylation sites in Atac2 predicted by SUMOsp ver. 1.0 online service.

Position (amino acid)	Predicted SUMOylation sites in Atac2
305	Glu-Lys-Gly-Glu
408	Ile-Lys-Gln-Glu
749	Phe-Lys-Thr-Glu

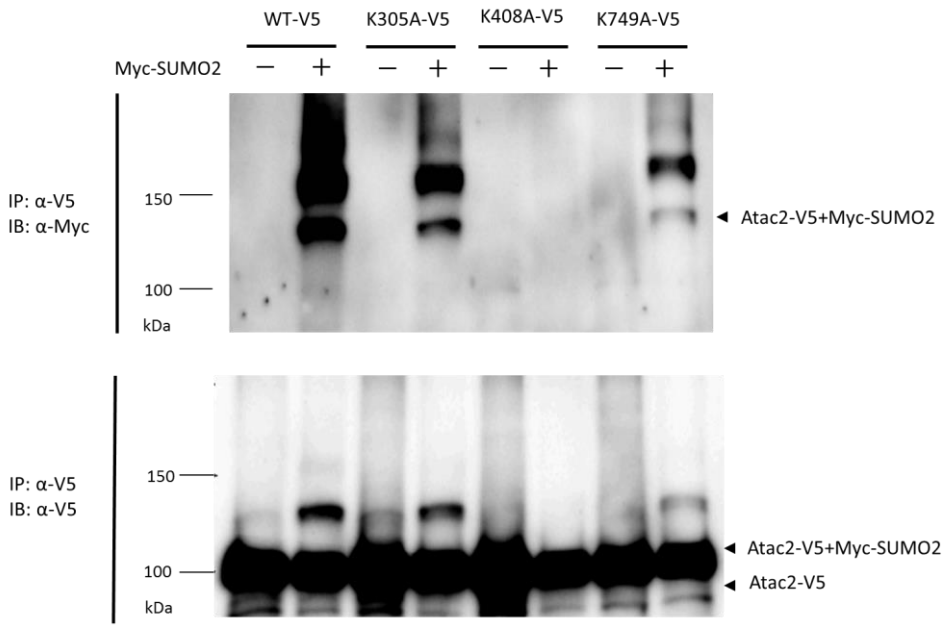

Fig.3-3 SUMO2 binds to K408 in Atac2.

NIH3T3 cells were transfected with V5-tagged wild type Atac2 or mutated Atac2 (K305A, K408A, and K749A) in the absence or presence of Myc-SUMO2 expression. Proteins were immunoprecipitated with anti-V5 antibodies, followed by Western blot with anti-Myc antibodies (Upper figure) and anti-V5 antibodies (Bottom figure).

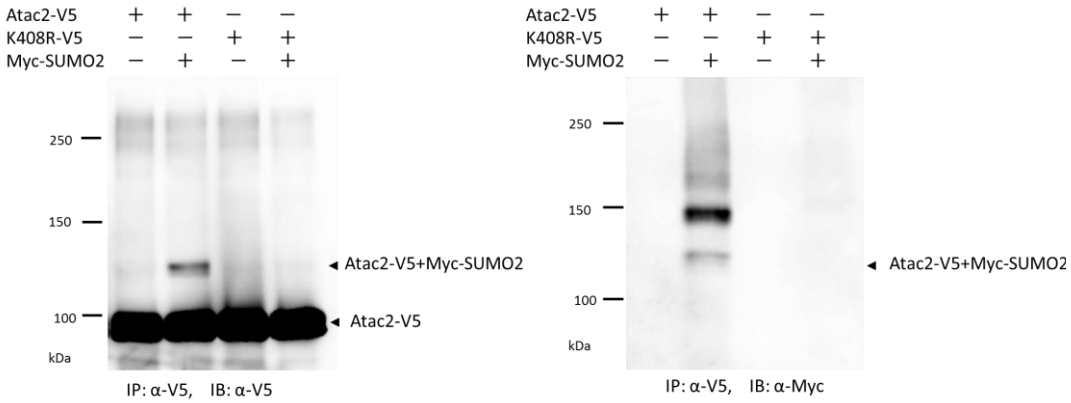

Fig.3-4. Assessment of non-SUMOylation of K408R mutant of Atac2.

NIH3T3 cells were transfected with V5-tagged wild type Atac2 or its K408R mutant in the absence or presence of Myc-SUMO2 co-expression. Proteins were immunoprecipitated with anti-V5 antibodies, and analyzed by Western blot with anti-V5 antibodies (Left) and anti-Myc antibodies (Right).

3-3-3. Confocal fluorescence microscopic analysis of Atac2.

SUMOylation sometimes changes subcellular localization of target proteins. To examine whether the localization of Atac2 alters upon the K408A mutation, NIH3T3 cells that expressed Venus-fused wild-type or K408A Atac2 were observed using a confocal fluorescence microscope (Fig. 3-5 A). Venus-Atac2 preferentially localized in the nucleus, which was consistent with a previous report[9]. Similarly, the Venus-K408A Atac2 mutant localized in a nucleus. This indicates that SUMOylation of Atac2 does not induce translocalization. Next, the location of SUMOylated Atac2 was visualized using the reconstitution technique of split Venus fragments. In VN-SUMO2 stable cell lines co-expressed with VC-Atac2, fluorescence of the reconstituted Venus was observed in the nucleus (Fig. 3-5 B). This indicates that the location of SUMOylated Atac2 is in a nucleus, which is an adequate result from the fact that the location of Atac2 itself is in a nucleus. Consequently, the SUMOylation of Atac2 does not affect its nuclear localization.

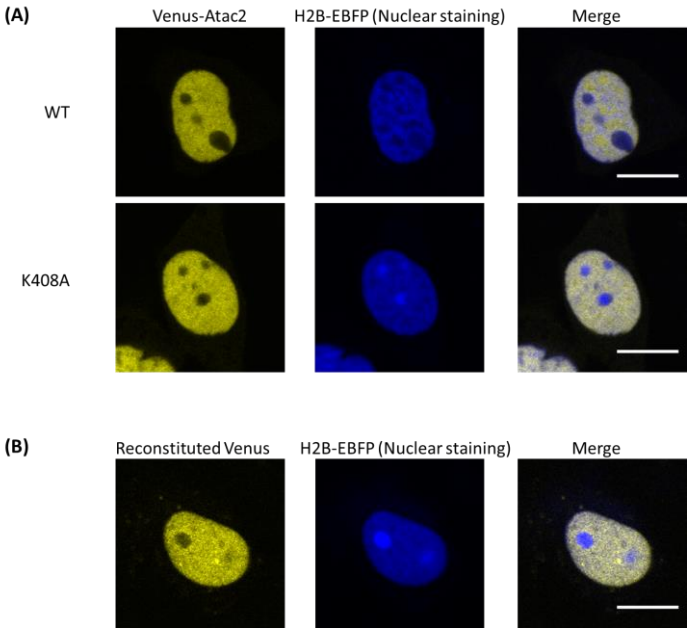

Fig. 3-5. Fluorescence microscopic analysis of localization of Atac2.
Fluorescence images were acquired from NIH3T3 cells expressing either Venus-fused wild-type or its K408A mutant with H2B-EBFP co-expression (A) and from VN-SUMO2 stable cell lines co-expressing VC-Atac2 and H2B-EBFP (B). Scale bar: 10 μm.

3-3-4. *In vitro* SUMOylation assay of Atac2.

Next, I investigated whether Atac2 proteins were modified by SUMO2 proteins under the *in vitro* condition. The wild type Atac2 and its K408R mutant was N-terminally fused with Flag epitope tag and purified from mammalian HEK293T cells. Recombinant proteins of N-terminally GST-tagged E1 (*Mus musculus* Aos1/Uba2), His-tagged E2 (*Xenopus* Ubc9), and His-tagged human SUMO2 were purified from *E. coli.*, respectively. Because it was reported that *in vitro* SUMOylation occurred without E3[10], E3 proteins were not used in this experiment. The purified Flag-tagged Atac2 proteins (wild type and K408R mutant) were reacted with ATP, E1, E2, and SUMO2 under various conditions where individual materials were absent or all the materials were present. The mixtures were analyzed by Western blotting with anti-Flag antibodies (Fig. 3-6). Only when Flag-Atac2 proteins were mixed with all the materials, the upshifted Atac2 proteins were detected around 120 kDa. On the other hand, the upshifted K408R Atac2 mutant proteins were not detected under the same condition. These results suggested that Atac2 was also modified by SUMO2 at K408 under an *in vitro* condition.

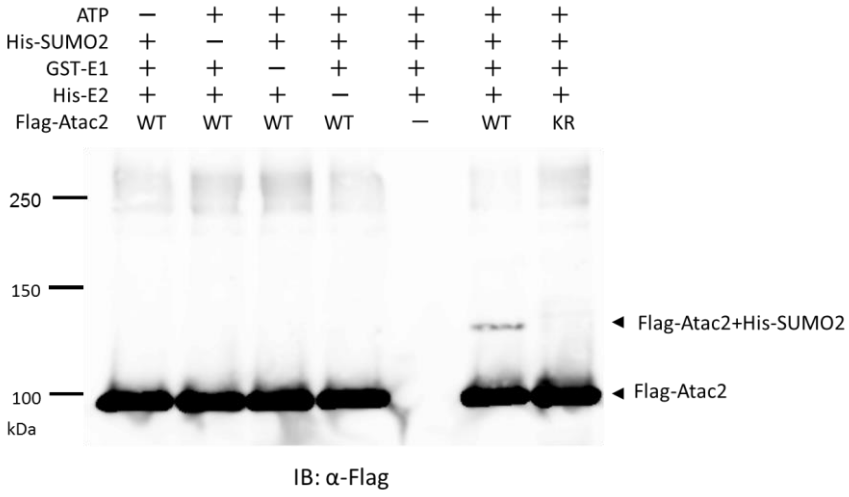

Fig.3-6 SUMOylation of Atac2 under *in vitro* condition.

Recombinant Flag-fused Atac2 proteins including wild type (WT) and the K408R (KR) were incubated with or without ATP, GST-tagged E1, His-tagged E2, and His-tagged SUMO2 as indicated. Reaction products were analyzed by Western blot with anti-Flag antibodies.

3-3-5. Examination of the origin of the unknown 150kDa Myc-SUMO2 band.

In the immunoprecipitation and Western blotting assay, an unknown Myc-SUMO2 band was detected over 150 kDa, where Atac2-V5 band was not detected. To clarify the origin of the unknown Myc-SUMO2 band, further analyses were conducted. First, Venus-fused Atac2-V5 (119 kDa) was subjected to the immunoprecipitation and Western blotting assay to discern which protein was the origin of the unknown 150 kDa band: modified Atac2, or another SUMOylated protein co-precipitated with Atac2 (Fig. 3-7). The Western blotting analysis showed that the Venus-Atac2-V5 modified by Myc-SUMO2 was detected at 150 kDa. This size was higher than that of Atac2-V5 modified by Myc-SUMO2 (120 kDa). The upshifted size (approx. 30 kDa) was consistent with the size of the fused Venus proteins (27 kDa). In a similar manner, the unknown 150 kDa Myc-SUMO2 band was upshifted to 180 kDa. This result indicated that the unknown Myc-SUMO2 band originated from the modification of Atac2. To detect the co-existence of Atac2-V5 with Myc-SUMO2 at 150 kDa by immunoblot with anti-V5 antibody, one more V5 epitope tag was N-terminally fused to Atac2-V5 (V5-Atac2-V5) to increase the number of bindable anti-V5 antibody and the immunoprecipitation and Western blotting analysis was conducted (Fig. 3-8). The immunoblotting with anti-V5 antibodies detected the second upshifted V5-Atac2-V5 band around 150 kDa. This ensured V5-Atac2-V5 proteins co-existed with Myc-SUMO2 proteins at 150 kDa. The reason Atac2-V5 was not detected with anti-V5 antibodies at the 150 kDa size was possibly due to the binding affinity of the anti-V5 antibodies. Collectively, the larger Myc-SUMO2 proteins detected at the 150 kDa originated from SUMOylated Atac2.

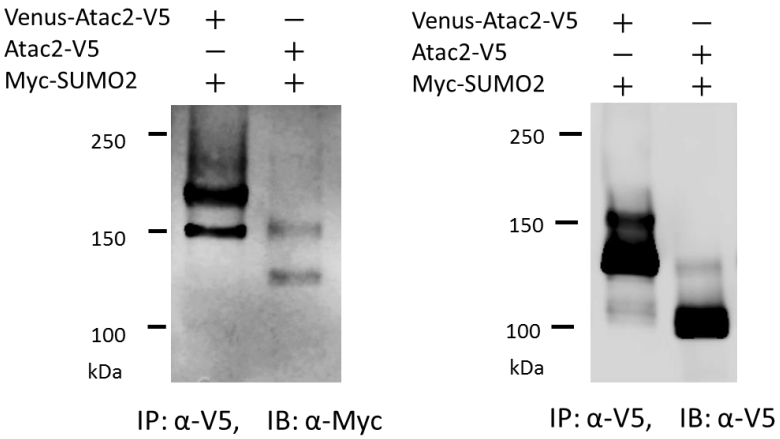

Fig. 3-7. The unknown 150 kDa Myc-SUMO2 band is originated from modified Atac2.

NIH3T3 cells were transfected with Venus-tagged Atac2-V5 or Atac2-V5 in the presence of Myc-SUMO2 co-expression. Immunoprecipitated proteins with anti-V5 antibodies from each cell were subjected to Western blotting with anti-Myc antibodies (Left) and anti-V5 antibodies (Right).

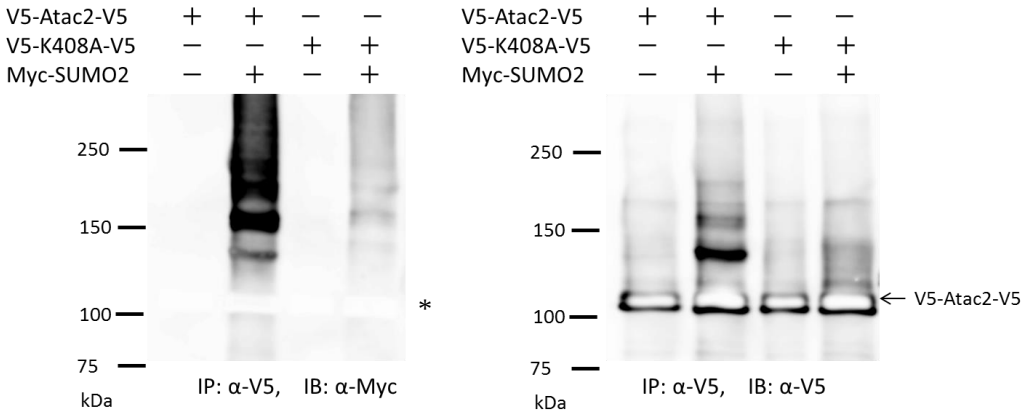

Fig.3-8. Detection of 150 kDa band with anti-V5 antibodies.

Immunoprecipitates from NIH3T3 cells expressing either double-V5-tagged Atac2 or its K408A mutant with or without Myc-SUMO2 co-expression were subjected to Western blotting with anti-Myc antibodies (Left) and anti-V5 antibodies (Right).

*After the membrane blotted with anti-V5 antibody was subjected to stripping to remove anti-V5 antibody, the same membrane was reblotted with anti-Myc antibody. Due to the stripping procedure, the locations where V5-Atac2-V5 was highly detected with anti-V5 antibody were blanked when blotted with anti-Myc antibody.

From the fact that the mutation in the 408 lysine residue in Atac2 resulted in detection of no Myc-SUMO2 bands including the 150 kDa band, it was highly possible that Atac2 was modified by polySUMO2 chain. It was also possible that SUMOylated Atac2 was modified by another protein modifier. To examine whether Atac2 was modified by polySUMO chain or not, K11 in SUMO2 was mutated to an arginine residue to prevent self-SUMOylation, and SUMOylation pattern of Atac2 was analyzed using the Myc-SUMO2 (K11R) mutant (Fig. 3-9). The result showed that the higher Myc-SUMO2 band around 150 kDa was still detected even though the SUMOylation site in SUMO2 was mutated. Furthermore, Myc-SUMO1, which doesn't form polySUMO chain, was used instead of Myc-SUMO2 in the immunoprecipitation and Western blotting assay. Beside the band corresponding to Atac2-V5 modified by single Myc-SUMO1 molecule, the higher Myc-SUMO1 band was detected at 150 kDa with anti-Myc antibodies. These results indicated another possibility that Atac2 was SUMOylated at not only a 408 lysine residue, but also another lysine residue. Subsequently, to discern whether the higher 150 kDa band was originated from Atac2 modified by at least two SUMO molecules, Venus-tagged Myc-SUMO2 (Venus-Myc-SUMO2) was used in the immunoprecipitation and Western blotting assay (Fig. 3-10 A). The band of Atac2-V5 modified by Myc-SUMO2 (120 kDa) was upshifted to approximately 170 kDa when Atac2-V5 was modified by Venus-Myc-SUMO2. In contrast, the higher 150 kDa Myc-SUMO2 band was upshifted to approximately 250 kDa. The upshifted size in case of using Venus-Myc-SUMO2 was larger than that in case of using Myc-SUMO2. This result suggests that one more SUMO molecule attached to Atac2. Similarly, when Venus-tagged Myc-SUMO1 (Venus-Myc-SUMO1) was used, the same SUMOylation pattern was detected (Fig. 3-10 B). Taken all results together, it was concluded that at least two lysine residues including K408 in Atac2 were targeted for SUMO modification.

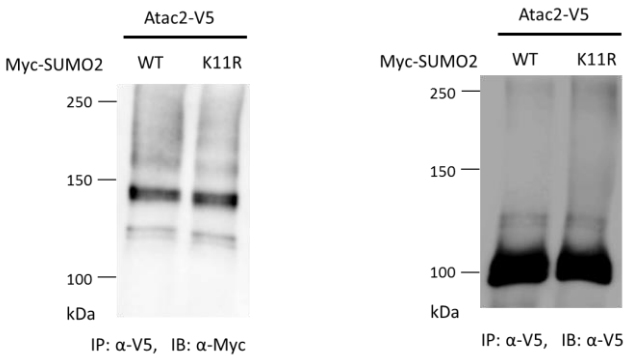

Fig. 3-9. SUMOylation assay of Atac2 using K11R SUMO2 mutant.

NIH3T3 cells were transfected with either Myc-SUMO2 wild type or Myc-SUMO2 (K11R) mutant with Atac2-V5 co-expression. Proteins were immunoprecipitated with anti-V5 antibodies and subjected to Western blotting with anti-Myc antibodies (Left) and anti-V5 antibodies (Right).

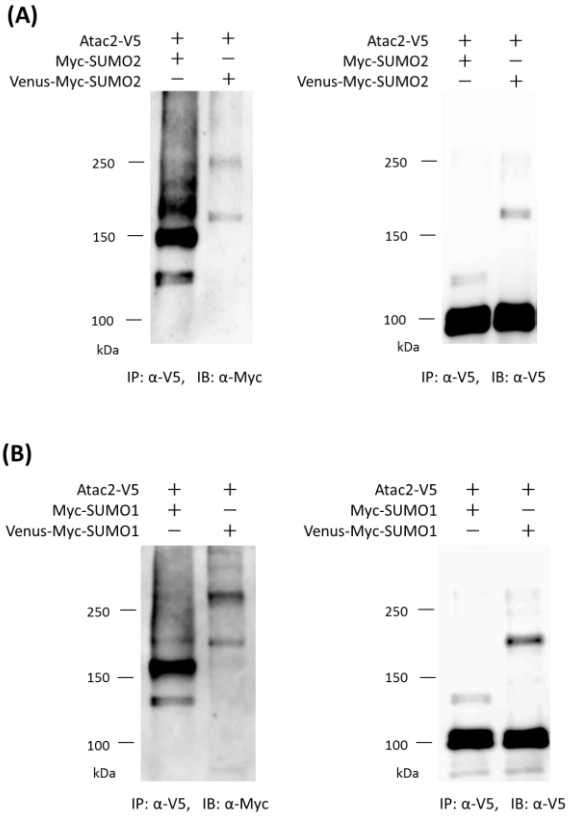

Fig. 3-10. Analysis of SUMOylation pattern of Atac2 using either Venus-fused Myc-SUMO2 or Venus-fused Myc-SUMO1.

NIH3T3 ells were transfected with Atac2-V5 co-expressed with either Venus-fused Myc-SUMO2 or Myc-SUMO2 (A), or Venus-fused Myc-SUMO1 or Myc-SUMO1 (B). The immunoprecipitated proteins with anti-V5 antibodies were resolved by SDS-PAGE and analyzed by Western blotting with anti-Myc antibodies (Left) and anti-V5 antibodies (Right).

3-4. Discussion

Using the present screening method, 36 kinds of proteins were identified as putative SUMOylated proteins. The immunoprecipitation and Western blotting assays to assess SUMOylation of these candidate proteins revealed that Atac2 was modified by SUMO2 at a 408 lysine residue. In addition, SUMOylation of Atac2 at a lysine 408 was also confirmed under *in vitro* condition. Unlike the result of i*n vitro* SUMOylation assay, the Western blotting analysis of the cell samples displayed another SUMOylation band around 150 kDa. From the analyses using Venus-Atac2, Myc-SUMO2 (K11R), Myc-SUMO1, and Venus-Myc-SUMO1/2, this band was considered as a form of Atac2 modified by two SUMO molecules. The reason the higher SUMO2 band was not detected in the *in vitro* assay is possibly the insufficient condition for further SUMOylation: insufficient amount of applied E1 and E2 enzymes; loss of other factors such as E3 enzymes. The analyses also indicated the possibility that one more lysine residue beside K408 in Atac2 was further modified by SUMO.

Atac2 is one of the components of a histone acetyltransferase complex, “ATAC”, which is essential for mammalian development[11,12]. Atac2 acts as a weak histone acetyltransferase and is necessary for the formation of the ATAC complex[12]. Fluorescence microscopic analysis revealed that SUMOylated Atac2 localized in the nucleus. This suggests that SUMOylation of Atac2 is related to its intra-nuclear activities. To clarify the function of Atac2 SUMOylation, further analyses are required in the future, such as examinations of the gene expression that is regulated by either Atac2 or ATAC complex and the effect on interaction with other components in the ATAC complex. The clarification of the functional effect of SUMOylation on Atac2 will provide a new insight into SUMOylation roles related to intranuclear events.

In the SUMOylation assays of candidate proteins, only Atac2 was clearly confirmed as SUMOylated. The reasons SUMOylation of other examined candidate proteins was not detected are possibly as follows. Firstly, the SUMOylation level of the candidate proteins might be too low to be

detected by Western blotting analysis. In general, less than 1% of the substrate proteins are SUMOylated at the steady state[13,14]. Due to the low SUMOylation level, the amount of the SUMOylated candidate proteins might be so small that only few amount of the antibodies specific to either SUMO or the candidates bind to them, which is possibly under the detection level of Western blotting analysis. In the case of Atac2, the higher 150 kDa SUMOylated Atac2 was difficult to be blotted with anti-V5 antibodies. Given that the binding affinity or qualities of antibodies influence on detection of SUMOylation, there is a possibility that Plscr3, immunoblotting of which detected multiple Myc-SUMO2 proteins but did not detect Plscr3-V5 proteins at the same positions, is also a novel SUMOylated protein. To prove SUMOylation by Western blotting, some improvements should be required: use of other antibodies; co-overexpression of SUMOylation enzymes to promote SUMOylation; scale-up of the cell samples. Another reason is that the candidate proteins are false positives. The fluorescence protein reconstitution has a potential to detect the interaction between the candidate proteins and the other SUMOylated proteins, because the reconstitution can occur if the split fluorescence protein fragments come close to each other. Indeed, Fth1 and Gsn, the SUMOylation protein candidates, were reported to interact with SUMOylated proteins, Daxx[15] and p53[16], respectively[17,18]. Similarly, the split fluorescent protein reconstitution might detect noncovalent interaction between the candidate proteins and SUMO2 via SUMO interacting motifs (SIMs)[19]. Indeed, among the screened protein candidates, 24 proteins such as Plxnb2, Ermp1, and Mrpl4 have putative SIMs. Further optimization such as replacement of the fluorescence protein fragments with different affinities would improve the reliability of the screening system.

Reference

1. Uchimura, Y., Nakamura, M., Sugasawa, K., Nakao, M. & Saitoh, H. Overproduction of eukaryotic SUMO-1- and SUMO-2-conjugated proteins in Escherichia coli. *Anal. Biochem.* **331,** 204–206 (2004).
2. Uchimura, Y., Nakao, M. & Saitoh, H. Generation of SUMO-1 modified proteins in E. coli: towards understanding the biochemistry/structural biology of the SUMO-1 pathway. *FEBS Lett.* **564,** 85–90 (2004).
3. Sarge, K. D. & Park-Sarge, O.-K. Detection of proteins sumoylated in vivo and in vitro. *Methods Mol. Biol.* **590,** 265–277 (2009).
4. Fukuda, I. *et al.* Ginkgolic Acid Inhibits Protein SUMOylation by Blocking Formation of the E1-SUMO Intermediate. *Chem. Biol.* **16,** 133–140 (2009).
5. Hietakangas, V. *et al.* Phosphorylation of serine 303 is a prerequisite for the stress-inducible SUMO modification of heat shock factor 1. *Mol. Cell. Biol.* **23,** 2953–2968 (2003).
6. Subramanian, L., Benson, M. D. & Iñiguez-Lluhí, J. A. A synergy control motif within the attenuator domain of CCAAT/enhancer-binding protein alpha inhibits transcriptional synergy through its PIASy-enhanced modification by SUMO-1 or SUMO-3. *J. Biol. Chem.* **278,** 9134–9141 (2003).
7. Eloranta, J. J. & Hurst, H. C. Transcription factor AP-2 interacts with the SUMO-conjugating enzyme UBC9 and is sumolated in vivo. *J. Biol. Chem.* **277,** 30798–30804 (2002).
8. Xue, Y., Zhou, F., Fu, C., Xu, Y. & Yao, X. SUMOsp: a web server for sumoylation site prediction. *Nucleic Acids Res.* **34,** W254–W257 (2006).
9. Ma, Y. *et al.* The CSRP2BP histone acetyltransferase drives smooth muscle gene expression. *Nucleic Acids Res.* **45,** 3046–3058 (2017).

10. Okuma, T., Honda, R., Ichikawa, G., Tsumagari, N. & Yasuda, H. In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. *Biochem. Biophys. Res. Commun.* **254,** 693–698 (1999).

11. Wang, Y.-L., Faiola, F., Xu, M., Pan, S. & Martinez, E. Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *J. Biol. Chem.* **283,** 33808–33815 (2008).

12. Guelman, S. *et al.* The Double-Histone-Acetyltransferase Complex ATAC Is Essential for Mammalian Development. *Mol. Cell. Biol.* **29,** 1176–1188 (2009).

13. Geiss-friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* **8,** 947–956 (2007).

14. Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **73,** 355–82 (2004).

15. Lin, D.-Y. *et al.* Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol. Cell* **24,** 341–354 (2006).

16. Stindt, M. H., Carter, S., Vigneron, A. M., Ryan, K. M. & Vousden, K. H. MDM2 promotes SUMO-2/3 modification of p53 to modulate transcriptional activity. *Cell cycle* **10,** 3176–3188 (2011).

17. Liu, F. *et al.* FTH1 binds to Daxx and inhibits Daxx-mediated cell apoptosis. *Mol. Biol. Rep.* **39,** 873–879 (2012).

18. An, J. *et al.* Gelsolin negatively regulates the activity of tumor suppressor p53 through their physical interaction in hepatocarcinoma HepG2 cells. *Biochem. Biophys. Res. Commun.* **412,** 44–49 (2011).

19. Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **281,** 16117–16127 (2006).

Chapter 4.

General conclusions

In this study, a new screening method for mammalian SUMOylated proteins was established based on reconstitution of split fluorescent protein fragments and FACS sorting. The present method has several advantages for the detection of mammalian SUMOylation. Conventionally, immunoprecipitation-based mass spectrometry (IP-MS) method[1,2] has been widely used for the screening of SUMOylated proteins. In the method, cells are lysed, and SUMOylated proteins are immunoprecipitated with antibody specific to a SUMO, followed by MS analysis. The cell lysis process potentially causes deSUMOylation by SUMO proteases and detectable SUMOylated proteins would be biased because highly and frequently SUMOylated proteins are preferentially collected from the cell lysates. In contrast, the present method uses reconstitution of split fluorescence protein fragments. The fluorescence protein reconstitution can detect SUMOylation under a live-cell condition. In addition, due to the irreversibility of the reconstitution reaction, the fluorescence signal is maintained even after deSUMOylation proceeds. Therefore, the present method has a potential to be detectable of infrequently-SUMOylated proteins. In the present system, a SUMO and a library protein were exogenously co-expressed in individual cells, and the fluorescence signals emitted from the reconstituted fluorescence proteins were detected in a single cell level by FACS. Thus, the detection of a weak SUMOylation signal in a cell is not disturbed by a strong SUMOylation signal in another cell, which is an advantage for the detection of low-level SUMOylation. As another method for detecting SUMOylated proteins, the yeast two-hybrid-based screening method has been deviced[3]. In the yeast two-hybrid method, translocation of both a SUMO and a target protein is required to induce expression of the reporter gene. SUMOylation of the proteins having difficulties to translocate into a nucleus cannot be properly assessed. In contrast, the fluorescence protein reconstitution in the present method does not require the translocation of a SUMO and a target protein. Thus, the present method does not limit the location of protein SUMOylation. In addition, the method using yeast cells has several difficulties in investigating

mammalian SUMOylation because the complexities of the mammalian SUMOylation system, including the specificity of SUMO isoforms and the SUMOylation patterns upon mammalian cell types, cannot be completely reflected in yeast cells. Therefore, the present method using mammalian cells is proper for exploring mammalian SUMOylated proteins. By using the present method, 2 SUMOylated proteins and 36 SUMOylated protein candidates with different subcellular locations and functions were successfully detected. Of the protein candidates, it was found that Atac2, a histone acetyltransferase[4,5], was modified by SUMO2 at a lysine 408 and localized in a nucleus, indicative of a new intra-nuclear regulatory role of SUMOylation.

The present method has a potential for further applications: screening of SUMOylated proteins modified by other SUMO isoforms; assessment of detectable SUMOylated proteins in individual mammalian cell-types; examination of detectable SUMOylated proteins under the condition that the extracellular stresses[6] are provided. Owing to such advantages in methodological analyses, the present method could be prospective for obtaining new insights into the important roles of SUMOylation in various biological contexts.

References

1. Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K. & White, M. A. Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J. Biol. Chem.* **279,** 20999–21002 (2004).
2. Tirard, M. *et al.* In vivo localization and identification of SUMOylated proteins in the brain of His6-HA-SUMO1 knock-in mice. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21122–21127 (2012).
3. Hannich, J. T. *et al.* Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. *J. Biol. Chem.* **280,** 4102–4110 (2005).
4. Wang, Y.-L., Faiola, F., Xu, M., Pan, S. & Martinez, E. Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *J. Biol. Chem.* **283,** 33808–33815 (2008).
5. Guelman, S. *et al.* The Double-Histone-Acetyltransferase Complex ATAC Is Essential for Mammalian Development. *Mol. Cell. Biol.* **29,** 1176–1188 (2009).
6. Tempé, D., Piechaczyk, M. & Bossis, G. SUMO under stress. *Biochem. Soc. Trans.* **36,** 874–878 (2008).

Appendix

SUMO consensus recognition sites, non-SUMO consensus recognition sites, and SUMO-interacting motifs (SIMs) in the proteins detected by the present method were computationally predicted from their amino acid sequences by using the web server GPS-SUMO 2.0 (Supplementary Table 1).

Supplementary Table 1. SUMO consensus recognition sites, non-SUMO consensus recognition sites, and SUMO-interacting motifs (SIMs) in the SUMOylated protein candidates predicted by GPS-SUMO 2.0, an updated version of SUMOsp ver.1.0.

Protein	SUMO consensus recognition sites	SUMO non-consensus recognition sites	SIMs
Rpl37a		V**K**SA(79-82), L**K**EL(86-89), L**K**DQ(89-92)	
Lmna	L**K**EE(88-91), A**K**LD(157-160)	A**K**LR(206-209), R**K**LE(307-310)	VVTIW(382-386)
Rps9	M**K**LD(92-95), L**K**IE(100-103)	N**K**RE(39-42)	
Rpl32			
Eif3e			IIDLF(232-236)
Gsn		E**K**FD(72-75), E**K**TE(717-720)	ILDLG(215-219), ITVVR(742-746)
Stx12		T**K**QD(52-55), S**K***(273, 274, stop codon)	LELIK(178-182), ILDVN(195-199), ILVLV(253-257), VVVIW(266-270)
Bgn		Y**K**K*(367-369, stop codon), K**K***(368, 369, stop codon)	LVLVN(120-124), LVEIP(151-155), LVELR(162-166)

Drosha	F**K**GE(1303-1306)	E**K**EA(388-391), D**K**LD(445-448), T**K**LD(470-473), R**K**YR(1345-1348), I**K**K*(1371-1373, stop codon), K**K***(1372, 1373, stop codon)	ILELY(655-659), LINIM(944-948)
Uqcrh	P**K**EE(16-19)	R**K**ML(7-10), F**K**NL(85-88), L**K***(88, 89, stop codon)	
Plxnb2	I**K**QD(442-445), V**K**AD(833-836), P**K**PD(1512-1515)	G**K**LD(1304-1307), N**K**VT(1837-1840)	IIVVS(1216-1220), LSVIA(1716-1720)
Rpl18a	M**K**VE(127-130)	T**K**RP(169-172)	
Atac2	L**K**GD(96-99), I**K**QE(407-410), F**K**TE(748-751)	E**K**GE(304-307), E**K**PD(416-419), C**K**HA(769-772)	LLIVE(37-41), IITVE(236-240), VVVLY(673-677)
Ermp1	V**K**LE(182-185), V**K**RD(697-700)		ILAVL(369-373), LLVIA(401-405), IIAVF(464-468), LLVCS(658-662)
Mrpl4			LDIVH(93-97), VLLVD(216-220)
Tmsb4x		D**K**PD(3-6), E**K**FD(11-14), K**K**TE(19-22), E**K**QA(38-41)	
Rpsa	M**K**EE(10-13)	E**K**EE(211-214)	IYIIN(46-50), IVAIE(65-69), VSVIS(74-78), LLVVT(121-125)
Lgals3			LITIM(145-149), IVLDF(159-163), VIVCN(184-188)
Pcolce		G**K**FD(202-205), R**K**CP(458-461)	VIMLT(196-200)
Tuba1b	V**K**CD(303-306)	E**K**DY(429-432)	IIDLV(114-118)

Pbrm1	L**K**ME(101-104), L**K**RE(934-937), F**K**SD(1033-1036), A**K**KE(1387-1390)	Q**K**GE(153-156), K**K**AE(283-286), K**K**KE(483-486), R**K**EL(637-640), E**K**EK(918-921), E**K**KE(920-923), E**K**LP(923-926), E**K**RE(939-942), E**K**SE(944-947), D**K**GD(1120-1123), A**K**FA(1329-1332), K**K**AE(1439-1442)	IILEP(576-580), IVCIE(992-996), VGVLG(1576-1580)
Myof	F**K**DE(28-31), L**K**RE(881-884), F**K**LE(1079-1082), P**K**EE(1351-1354), A**K**KE(1851-1854), P**K**LD(1981-1984)	K**K**VD(36-39), E**K**RD(332-335), T**K**ND(449-452), K**K**LE(539-542), N**K**FD(591-594), R**K**KD(1014-1017), E**K**GP(1095-1098), G**K**SD(1493-1496)	VSVIF(24-28), LVIVV(65-69), IDLVI(118-122), IISIR(260-264), LLVVE(549-553), VVTLT(627-631), LLEIE(738-742), IIIWM(762-766), IEILA(1293-1297), VIEIE(1416-1420), LEVLN(1957-1961)
Dynlrb1		K**K**NE(74-77)	IIVVN(20-24), LIVI**Q**(88-92)
Fam63b	W**K**EE(228-231)	D**K**EK(588-591), E**K**EK(590-593), E**K**EK(592-595), E**K**NS(594-597)	VIL(599-601)
Taz			
Rps3a	P**K**FE(221-224)	S**K**KD(26-29)	
Myl9		D**K**ED(50-53), L**K**HG(163-166), D**K**DD(169-172)	
Rpl6		T**K**VE(85-88)	LIILT(158-162), LLLVT(180-184)

Narf	MKCE(1-4)	S**K**TD(15-18), E**K**GE(42-45), K**K**LE(237-240), I**K**W*(460-462, stop codon)	VEVLA(375-379)
Arpc1b		V**K**SL(359-362), L**K**DL(366-369), L**K**IK(369-372), I**K***(371, 372, stop codon)	LVILR(90-94), VISIC(119-123), ISVLS(332-336)
Psmb4	V**K**FD(59-62)		
Polrld		R**K**AI(9-12), D**K**CD(117-120)	
Rpl10			VIRIN(96-100)
Fth1		D**K**ND(124-127)	
Anxa5	L**K**SE(67-70), I**K**GD(298-301)	G**K**FE(73-76), L**K**WG(183-186), Y**K**KA(306-309), K**K**AL(307-310)	LVVLL(151-155), IRVVV(268-272), LLLLC(310-314)
Plscr3		T**K**DE(227-230), E**K**RG(284-287)	
Wisp2			
Cops7a		A**K**IW(268-271)	LVIEA(131-135)

* indicates stop codon.

The character "**K**" with bold highlight indicates the predicted SUMOylation site.

The numerals in the table indicate the position of amino acids.